NATIONAL CONVENTION TRANSACTIONS 1959

THIRTEENTH ANNUAL CONVENTION

AMERICAN SOCIETY FOR QUALITY CONTROL

NATIONAL CONVENTION TRANSACTIONS 1959

THIRTEENTH ANNUAL CONVENTION

AMERICAN SOCIETY FOR QUALITY CONTROL

MAY 25, 26, 27, 1959

CLEVELAND, OHIO

PRICE

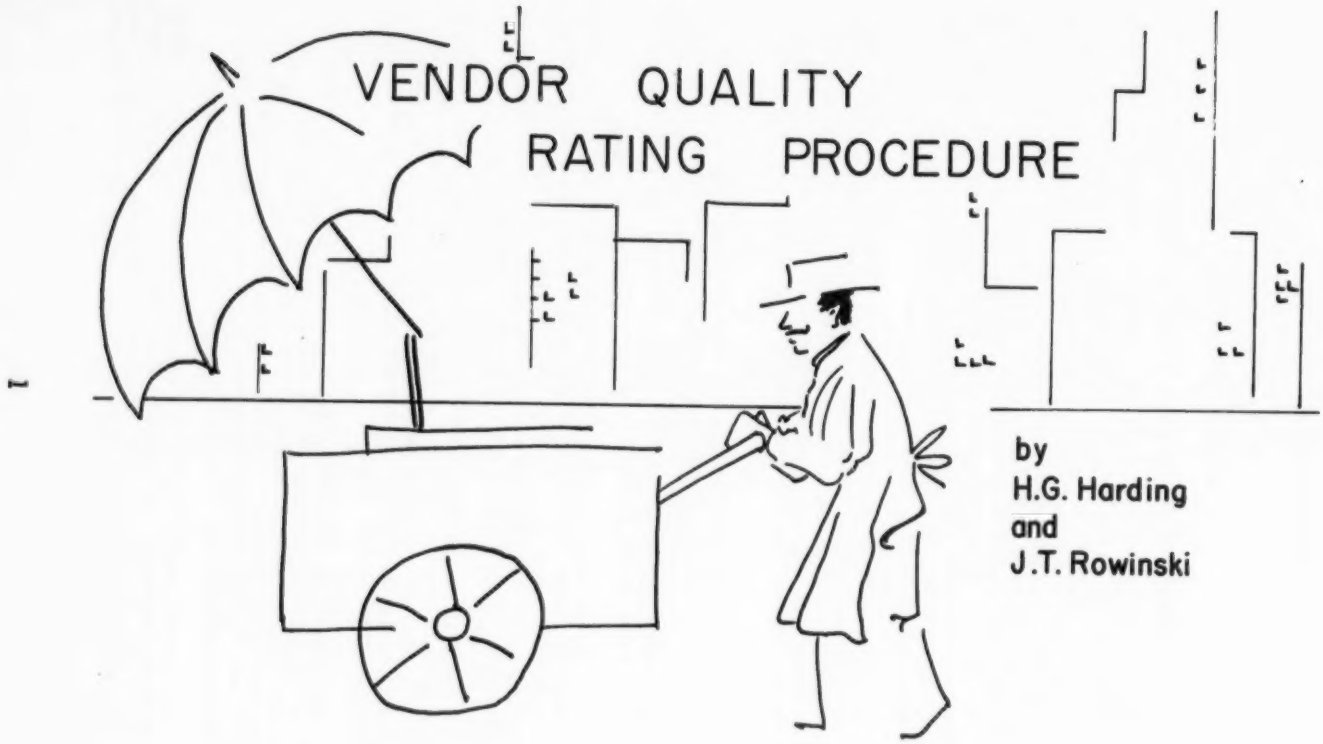
$3.00 at the Convention

$3.50 if Ordered by Mail

Copyright 1959, By
AMERICAN SOCIETY FOR QUALITY CONTROL, INC.
Room 6197
Plankinton Building
161 West Wisconsin Avenue
Milwaukee 3, Wisconsin

Printed in the United States of America By
UNIVERSAL LITHOGRAPHERS, INC.
Baltimore, Md.

VENDOR QUALITY RATING PROCEDURE



by
H.G. Harding
and
J.T. Rowinski

VENDOR QUALITY RATING PROCEDURE

Harry G. Harding
and
Joseph T. Rowinski

International Business Machines Corporation

Quality rating of vendor-produced components, so vital to any efficient industry, has been, for some time, an area of persistent confusion. Vendors are often at loss to discover why one shipment of their product is accepted and another one is rejected. In some instances, the vendor may feel that acceptance is based on factors not pertinent to the successful performance of his product; at other times, he may feel that a completely arbitrary set of standards has been applied, subject to the whims of the inspecting department. Confusion and misunderstanding such as this are conditions that IBM has steadily sought to avoid. The IBM Military Products Rating System is our attempt to clear up all areas of acceptance standards from the vendor's viewpoint.

Our repeated use of the word "vendor" may conjure up such a sight as (frontispiece) our old friend the peanut vendor—or is it horse chestnuts? However, without belittling the pushcart peddler, we have no desire to infer that IBM buys from a moving cart. Our vendor is a business man who wishes, for his own interests, to give us the best product possible.

For this reason, the IBM Military Products Rating System, which must monitor the quality of components for the gigantic SAGE computer, with its 58,500 vacuum tubes, its 170,000 diodes, its 400,000 resistors, and countless other parts, has been designed to provide each vendor with a monthly rating of his product.

This rating, based upon established acceptable quality levels (AQL), will show the vendor into which of two groups (acceptable and unacceptable) his product has fallen within the preceding 30 days. It will also inform him as to the reason or reasons for any rejections so that he may make any required corrections before his next shipment.

In addition to the vendor monthly report, a comparison report showing the product quality trend for a 3-month period is given to each vendor. This constant reporting on deviations (from what is acceptable quality in either direction) is intended as an aid to all vendors supplying components to IBM Kingston.

Underlining the important points, I should like to take a few minutes to describe the workings of our Vendor Rating System and at the same time (fig. 1) call attention to the form used internally in preparing individual ratings. A different form, named "Purchased Product Quality and Delivery Conformance Report," is used to notify the vendor of his rating, based upon the results of our internal report.

The internal report takes into account the number of lots received, the quantity inspected, the quantity found defective, the AQL assigned, and the number of lots rejected. Each vendor is given a numerical rating with a potential of 100. There are three rating levels: 100 through

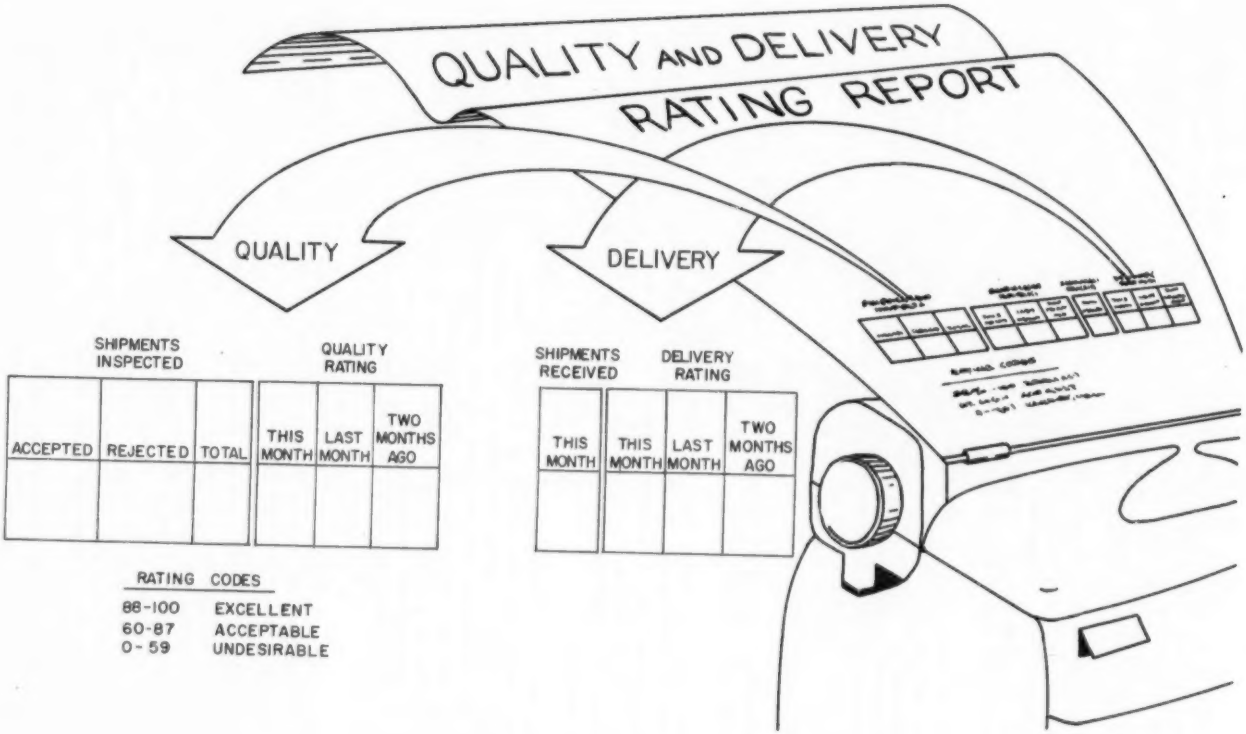


Figure I. Internal Quality Rating Report

88(excellent), 87 through 60 (acceptable), and 59 through 0(undesirable). Our internal form also includes the delivery rating, showing the delivery trend for a 3-month period. Based on his numerical rating, the vendor is put into an "acceptable" or "unacceptable" category and is so notified on the Purchased Product Quality Conformance Report.

We have prepared a graph (fig. 2) showing vendor ratings for each month within a 7-month period ending June 30, 1957. You may note that on this graph about 30 percent of the vendors were in the 100 category and about 30 percent were in the undesirable class. Breakdowns were arrived at by the IBM Kingston Quality Control Groups.

The ultimate goal of vendor rating that would be represented on this graph is that of having the majority of the vendors in the "excellent" and "acceptable" categories; namely, those areas shown on figure 2 by the gray and green areas, respectively. The reason for this is apparent. It is considered of prime importance to see that the vendor who is in the "undesirable" category move up into the "acceptable" category. In order to do this, a careful analysis of the vendor's quality system is necessary, and engineering help is required.

Receiving inspection makes available quantities of information which are sorted and culled for data that is applicable to our needs (fig. 3).

Our first criterion used was the following: If in the sample taken there were no defectives found, that particular vendor deserved the highest rating possible, a necessary expedient for good vendor-consumer relations. It is virtually impossible to explain to a vendor why he did not receive the highest rating possible if there were no defectives found in his shipment.

The second criterion used was that of the actual percentage of defectives found, for the period of a month, having a specific AQL for an individual vendor. Since the percent of defectives found in our samples is the best estimate that we have of the process capabilities of a vendor, and since the AQL designated is that particular percent defective generally accepted 95 percent of the time, it was decided that these two pieces of information would be used as a factor. We divided the actual percent defective by the assigned AQL and thus acquired a factor available for use.

The third criterion used was that of grading by quality, using the same terminology used in our school systems; i.e., 100 percent designates perfection, an average grade is 75 percent etc. With these three points established, part of our system was now designed. At the outset, all vendors receive 100 points. From this rating some value is subtracted, which is a function of the factor that is found by means of dividing the percent defective by the AQL. In order to have this consistent with our consideration of a norm, we decided that if this factor were equal to one (that is, if the percent defective found was exactly equal to the AQL), this was an average vendor. All that was necessary, then, was to determine what point was to be considered as norm and then to multiply this AQL factor by a constant which would give the average grade.

This now supplied us with one point from which to work. However, since an average rating could not be considered at one point only, we

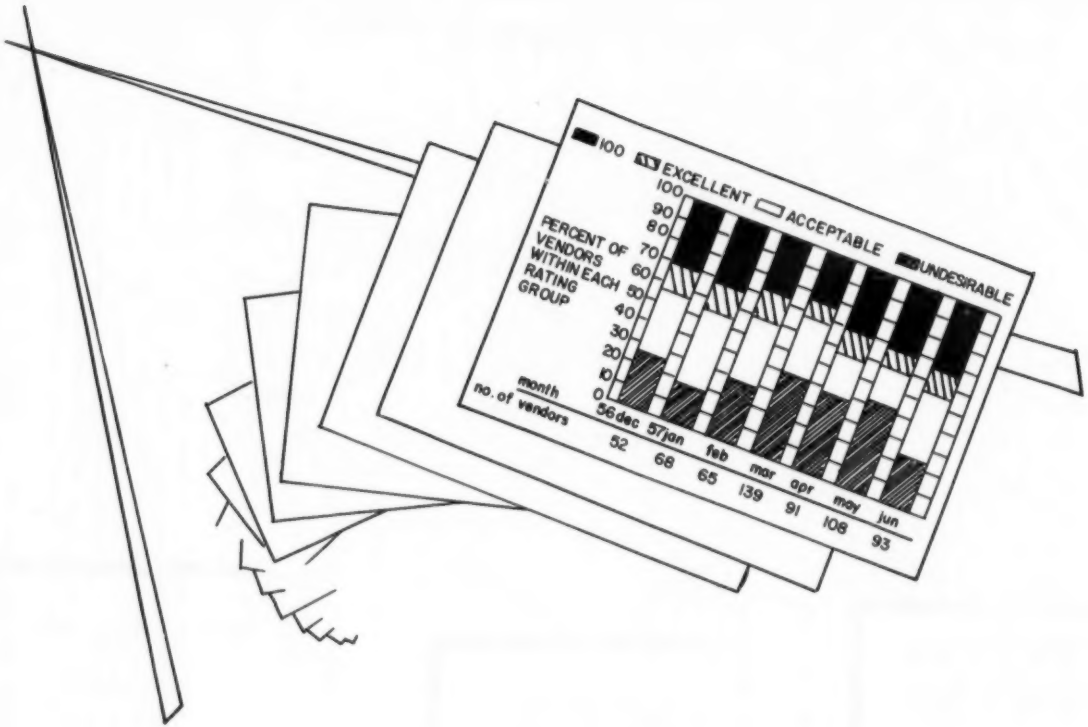


Figure 2. Vendor Quality Rating

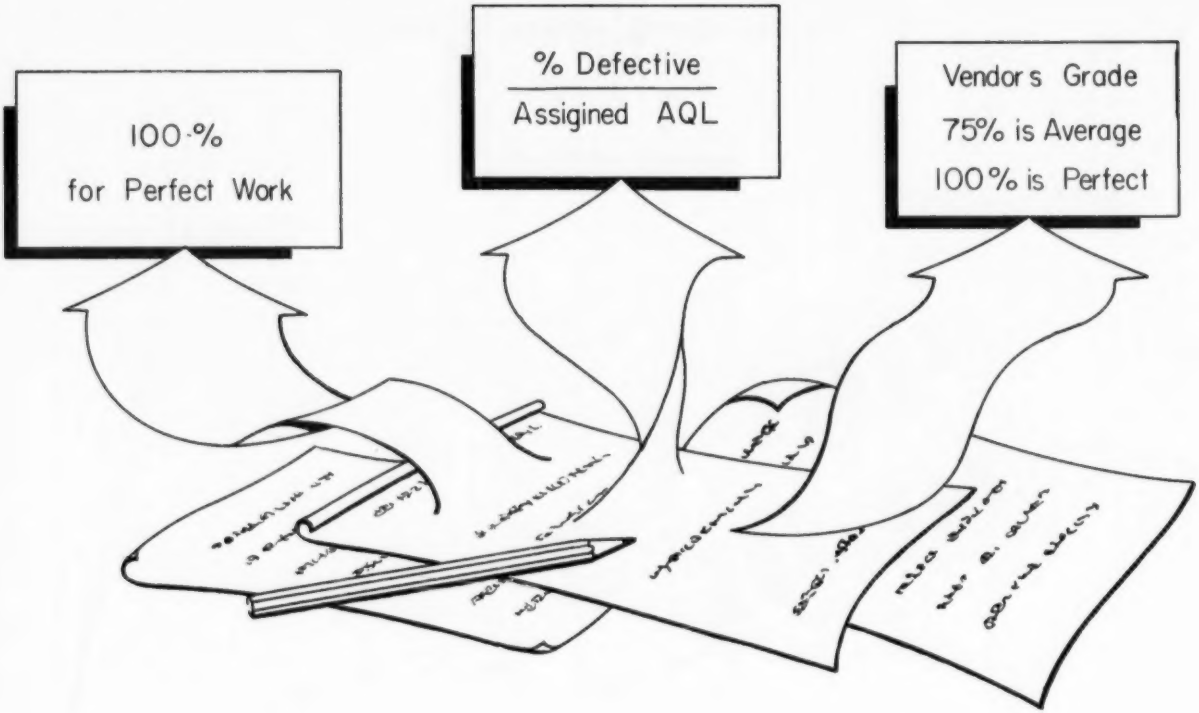


Figure 3. Mathmatical Derivation

decided that a spread for average was needed. This was accomplished by determining the AOQL's for each sample size. Then, by averaging these AOQL's we found that on the average the AOQL was 1 5 times the AQL.

Using this fact as a base, we were now able to determine our low limit (that point at which a vendor was to be considered an undesirable). Any vendor whose lots had a percent defective equal to or greater than 1.5 times the AQL assigned to him would now get an undesirable rating. By the same token, it was decided to equalize the range. In this way a vendor whose lots had a percent defective of one-half the AQL or better would get an excellent rating. All that was necessary to be done now was to calculate the range and, as we mentioned previously, the average would be 87-60.

Normally, this would suffice in a vendor rating procedure, but we had found that there were two more factors that had to be considered (fig. 4). These two factors were the number of lots received and the number of lots rejected.

A vendor who generally submits lots that are of excellent quality but who, every so often, submits a lot that has to be rejected is not necessarily a good vendor. For this reason, we decided that a further subtraction of points from the potential of 100 was necessary. This subtraction, we decided, was to be some function of the number of lots rejected, divided by the square root of the number of lots inspected.

Once again, all the constants that were needed to give an average rating of 87-60 were recalculated on this assumption. A vendor whose lots had a percent defective that was 1.5 times the AQL assigned and whose lots were accepted 95 percent of the time constituted vendor worst-quality level that we could consider as average. By the same token, a vendor whose lots had a percent defective of one-half the AQL assigned and whose lots were accepted 99 percent of the time became the vendor worst-quality level that we could consider as excellent. These conclusions gave us the constants necessary for the completion of the formula.

We feel that the merits of this system are apparent:

1. If no defectives are found, a rating of 100 is awarded.

2. This rating is influenced only by percent defective.

3. AQL factors, alone, are judged. The vendor can easily adjust his standards to raise his score.

4. The rating utilizes only the results from the samples which were selected, and by the constant factor mentioned. Here again, the system rates only the actual product. It is performed with a consistent accuracy and is set forth in terms that are easily understandable and clear to the vendor. These ratings are not influenced by any factors that the vendor cannot control. He can use them as accurate guides toward improving his product.

Inspection results can be punched on IBM cards, and the sample calculations can be swiftly completed by machine accounting. Only those factors which have a direct bearing on the quality of the product are

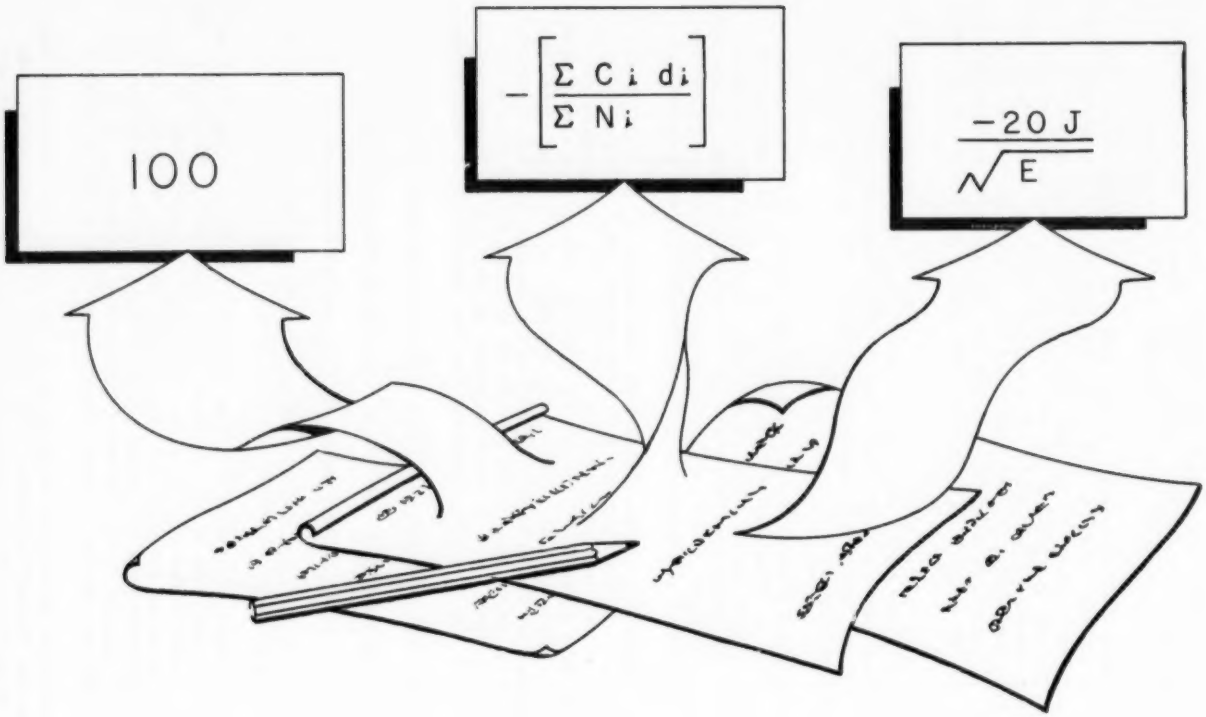


Figure 4. Mathmatical Derivation

included in the calculations.

These results get posted upon the internal rating report. With this rating accurately and promptly determined, all personnel who are concerned with vendor relationships now are able to exercise their functions of estimating purchasing and quality control with increased accuracy and efficiency. Based upon this rating, the vendor's copy of the Purchased Product Quality and Delivery Conformance Report is sent out.

If the vendor's shipment is unacceptable (in part or in total), a Rejected Purchase Report (RPR) is also prepared. This additional report not only informs the vendor that his merchandise is unacceptable, but it also gives the reasons. The report is numbered, and all subsequent conversation and correspondence with the vendor will refer to the RPR report number.

As the customer, IBM Kingston Military Products is more interested in the acceptance of products than in their rejection. For this reason it is felt that the system discussed here is the best method for securing such acceptance. The vendor is notified promptly of the acceptance or rejection of his product. By availing himself of the full advantages of this rating system he can improve his acceptance record. Thus, both IBM and the vendor achieve the same objective.

QUALITY CONTROL TECHNIQUES IN TIME STUDY

M. E. Mundel
M. E. Mundel & Associates

The task of managers is to design, predict the performance of and control an integrated, human-group activity, the related physical facilities and the inter-relationships between the two. To design or make detailed plans for an activity requires a dimension to work with, a dimension which permits the fitting together of diverse activities. The dimension must be a physical dimension so as to permit realistic control of the activity. The only dimension which fulfills these requirements, on a day-to-day basis, is some function of time. The measurement of the time that work should take, the appraisal of work in terms of time, is a basic requirement for effective management.

The importance of time measurement was recognized at the very inception of modern management. Taylor, in his historic paper, "Shop Management" in 1903, described the timing of work (time study) in order to set a realistic goal of performance for the workers. The appraisal of work in terms of time, time study, has many uses in modern management. It may be used:

"1. To set schedules.
2. To determine supervisory objectives.
3. To determine operating effectiveness.
4. To set labor standards.
5. To determine the number of machines a person may run.
6. To balance the work of crews, coordinate or in sequence.
7. To compare methods.
8. To determine standard costs.
9. To determine equipment or labor requirements.
10. To provide a basis for incentive wages."[1]

Time study or work measurement, as it is commonly called, has been a difficult industrial problem. Early techniques had many inaccuracies. Inasmuch as time standards, (the measurements resulting from time studies) often affect wages, often time studies have been the subject of grievances. They have been attacked categorically by various union as "unscientific," or "not even measurement."

To fully understand the problem involved, it must be realized that the dimension actually used is more than just "time." A time standard, to be applicable for the uses previously listed must represent the time that should be taken, using a specified method and with specified working conditions, by some particular type of worker, exerting himself to some specified degree.

This is indeed a laborious description of a unit of measurement, but all of the items listed are necessary. It is not the purpose of this paper to discuss time study or the definition of standard time as such, but the definition of the unit of measurement is necessary if we are to intelligently discuss quality control aids for applying this unit of measurement. To briefly summarize the entire problem, there are five major steps in time study as follows:

1. Setting up the unit of measurement.
2. Recording the method and conditions of work.
3. Observing an actual performance.
4. Relating the performance observed to the concept implied in the unit of measurement.
5. Adding allowances for various activities such as personal time.

We have already indicated some of the problems of setting up the concept of the unit of measurement. The solution of this problem is not primarily a statistical problem and I will not discuss it further in this paper; it can be solved in a reasonable manner.

The second step is a mechanical one and likewise, is not within the scope of this paper.

The third step, that of observing an actual performance, is a statistical problem. If we are going to manipulate a value representing an observed performance, we need some assurance that the figure we have is actually representative, at least reasonably so. Further, we need some way of quickly determining how much data to gather, while we are gathering data; to go back later for more data would certainly increase our chances of getting data from two levels of performance.

If we set our risk limits as being "enough data so we have 95 chances out of 100 that additional data would not change our average, observed value by more than 5% of the average," we will be well within limits which are suitable (sociologically) for the measurement.

Our measurement should be such that it may be made rapidly and without much computation. What we are seeking is a condition where:

$$2\sigma_{\bar{x}} \leq .05\bar{X}$$

If we estimate σ as $\bar{R}/d_2$ and $\bar{X}$ as the highest value (H) plus the lowest value (L) divided by two, we may write:

$$\frac{2(H-L)}{d_2\sqrt{N}} = .05\left(\frac{H+L}{2}\right)$$

And for given sample sizes, we can express N as a function of $\frac{H-L}{H+L}$ and create a simple table for N as in Table I, Figure 1.

The time study man, when making his observations, can take 10 readings, compute the H and L value for the ten readings or for each of the two samples of 5, and then quickly compute the $\frac{H-L}{H+L}$. Using the table as a guide he can then determine the minimum number of readings he should make of the performance he is observing.

TABLE I[2]

Number of Readings Required for ± 5 Per Cent; 95/100 Probability*

$\frac{H-L}{H+L}$	Data from sample of		$\frac{H-L}{H+L}$	Data from sample of		$\frac{H-L}{H+L}$	Data from sample of	
	5	10		5	10		5	10
.05	3	1	.21	52	30	.36	154	88
.06	4	2	.22	57	33	.37	162	93
.07	6	3	.23	63	36	.38	171	98
.08	8	4	.24	68	39	.39	180	103
.09	10	5	.25	74	42	.40	190	108
.10	12	7	.26	80	46	.41	200	114
.11	14	8	.27	86	49	.42	210	120
.12	17	10	.28	93	53	.43	220	126
.13	20	11	.29	100	57	.44	230	132
.14	23	13	.30	107	61	.45	240	138
.15	27	15	.31	114	65	.46	250	144
.16	30	17	.32	121	69	.47	262	150
.17	34	20	.33	129	74	.48	273	156
.18	38	22	.34	137	78	.49	285	163
.19	43	24	.35	145	83	.50	296	170
.20	47	27						

* For ± 10 percent, 95/100 probability, divide answers by 4.

Figure 1. Sampling table for time studies.

It is true, if we take only one sample, from which to form our estimate of $\bar{R}$, that we are using R as if it were $\bar{R}$, or using only two values of R to obtain $\bar{R}$, admittedly not a fully valid procedure. Also, we are using $\frac{H+L}{2}$ as equal to $\bar{X}$ which may not be accurate either.

There are however, some related points work considering. For one thing, performances including foreign or unwarrented motions are usually deleted, so extreme values of R are not obtained.[3] Secondly, the observer is only seeking a guide. With this guide he may then continue to observe in order to obtain the indicated number of readings and then test this larger sample using a true $\bar{R}$ and the actual $\bar{X}$ to determine if the final number of readings do indeed meet the criterion which was set up. It should also be noted that in some cases, the indicated number of readings is rather large. Because of shop exigencies, the time study man may not be able to obtain this many. This should not mean that the obtained data cannot be used, but rather that it should be used with care.

In time study work, the statistical measures should not be used as a rule, but rather as a guide. This principle has some interesting sidelights. Of late, some unions have been seeking to insert in labor contracts, a statement that "a sufficient number of observations should be made to provide a representative average of the performance observed." As a working guide this can be a suitable statement, but the Quality Control group should inform its management how insoluable such a re-

quirement could become when a difficult grievance arises. Representative may only be stated as being between limits with a certain probability. If certainty is required, then no time study is valid.

The fourth step of time study, that of "relating the performance observed to the concept implied in the unit of measurement," has been one of the main sources of error and is frequently the subject of attack. To perform this step, the time study man, through a judgement process, must compare the performance observed, to a concept of performance implied in the full statement of the unit of measurement (as previously given), and express the relationship quantitatively, usually as a percentage. All too often, both the managers who must enforce the standards set, and the supervisors of those who make time studies, know nothing about the confidence limits of such judgements as rendered by their time study men. Complex research techniques have their place in a study of such judgements, or ratings as they are called, but a quick and easy analytical technique is more readily used and understood. This is the same as with S.Q.C. techniques applied to product measurements.

Figure 2 is a simple analysis, using S.Q.C. techniques, of a series of ratings made by eight experienced time study men in a plant.

They were shown a film consisting of ten sequences of repetitions of the same job, each sequence being of a repetitive performance at a different speed of work. Their ratings (evaluations of the paces of work) appear in the upper half of the sheet.[4] The column labeled "A" is an arbitrary series of ratios giving the actual relative speeds of each scene. The K in the upper right hand corner is the value by which these arbitrary values must be multiplied if the group values, when plotted against the arbitrary values, are to have the least squared deviations about a 45° line, a line all points should be on if a consistent concept of normal was maintained and consistently applied. We will have more to say about this value K, later.

The A values are multiplied by K to obtain the KA column. Using these KA values as the best available approximation of the correct rating, the bottom half of the table is computed as percentages of these values. In this manner, the bottom table becomes a set of comparable values and the normal S.Q.C. techniques may be applied as indicated at the bottom of the sheet. These values may be plotted on control charts by men, by work speeds and so forth, permitting the supervisor of these men to determine the confidence limit of ratings and guide himself accordingly. Such an analysis also permits him to critique the rating performance of each man and to assist the men toward more accurate and consistent ratings. An analysis of this type is made each month at the plant from which these data were taken.

A control chart for one of the time study men is shown in Figure 3.

If films are to be used more than once within a reasonable period of time, a very large number of films is required, because most time study men appear to have an unusual memory for ratings. In the plant from which these data were taken, a new series of ten films (different speeds of doing work), of a different job, are prepared each month.

RATING ANALYSIS CALCULATION

Organization Patrick Cudahy Inc
Operation Move Bags of Hard Flakes
Notes TRAINED T.E. MEN
Date 8/10/53

$K = \frac{\Sigma \bar{X} A}{\Sigma A^2} = \frac{23532}{45772} = .824$

RATINGS

File	MEN A	B	K	Be	Br	M	N	P								Sum	$\bar{X}$	A	KA
1	92	56	68	74	50	70	65	68								493	62	71	59
2	79	73	79	84	68	79	88	85								625	78	100	82
3	57	62	55	53	59	56	58	55								455	56	69	57
4	98	67	59	58	65	60	68	70								495	62	79	65
5	34	45	42	38	40	42	40	36								317	40	53	44
6	70	67	57	62	55	58	66	63								498	62.	75	62.
7	92	85	80	88	76	82	89	96								688	86	93	77
8	40	61	44	40	42	40	50	40								357	45	61	50
9	52	72	67	66	68	65	70	74								534	67	82	68
10	81	87	82	82	82	83	92	89								678	85	101	83

RATINGS CONVERTED TO PERCENT OF CORRECT

									R
1	71	95	105	126	85	119	110	105	98
2	90	89	90	103	83	97	108	104	25
3	90	109	97	93	104	98	102	97	19
4	79	103	91	89	100	92	105	108	34
5	77	102	96	86	91	96	91	82	25
6	113	108	92	100	89	99	106	102	24
7	120	111	104	114	93	107	105	125	21
8	80	122	88	80	84	80	100	80	42
9	77	106	99	97	100	96	103	109	29
10	98	105	99	99	99	100	111	107	13
$\bar{X}$	89	105	96	99	93	98	104	102	98 = $\bar{X}\bar{t}$ $\bar{R}f$ = 28
R	99	32	17	39	21	39	20	45	32 = $\bar{R}m$

ANALYSIS BY MEN

$\bar{R}m = 32$ $\bar{\bar{X}}m = 98$

$UCL\bar{X}_m = \bar{\bar{X}}_m + A_2\bar{R}m = 98 + .308(32) = 108$

$LCL\bar{X}m = \bar{\bar{X}}m - A_2\bar{R}m = - - - = 88$

$UCLRm = D_4\bar{R}m = 1.777 \times 32 = 57$

$LCLRm = D_3\bar{R}m = .223 \times 32 = 7$

ANALYSIS BY PACE

$\bar{R}f = 28$ $\bar{\bar{X}}f = 98$ $SD = \frac{\bar{R}f}{d_2} = \frac{28}{2.85} = 9.8$

$UCL\bar{X}f = \bar{\bar{X}}f + A_2\bar{R}f = 98 + .373(28) = 108$

$LCL\bar{X}f = \bar{\bar{X}}f - A_2\bar{R}f = - - - = 88$

$UCLRf = D_4\bar{R}f = 1.864 \times 28 = 52$

$LCLRf = D_3\bar{R}f = .136 \times 28 = 4$

Figure 2

M. E. Mundel 570115-5

RATING ANALYSIS CALCULATION

Organization Patrick Cudahy
Operation Mixer
Notes TRAINED MEN
Date 11/24/53

$K = \frac{\Sigma \bar{X} A}{\Sigma A^2} = \frac{23132}{39993} = .578$

RATINGS

File	MEN A	B	K	Be	Br	M	N	P								Sum	$\bar{X}$	A	KA
1	67	72	73		67	70	70	73	Move load blocks = 7							492	70	77	75
2	83	82	77		73	78	76	78	Lift boxes, = 10							547	78	74	72
3	46	34	43	Absent this test	40	44	46	48	Push real carcass = 3							301	43	47	45
4	90	101	96		108	97	95	106	Transfer bundles = 6							693	99	98	95
5	34	35	42		34	46	44	38	Split trailer = 3							273	39	37	36
6	43	44	48		34	53	53	47	Pile box on skid = 4							322	46	51	49
7	40	40	50		50	47	50	48	Push real carcass = 11							325	46	43	42
8	123	115	114		120	114	108	116	Pull hand truck = 5							810	116	139	135
9	68	74	73		64	70	68	68	Pile box on skid = 8							485	69	73	71
10	55	70	68		67	72	75	74	Split trailer = 7							481	69	59	57

RATINGS CONVERTED TO PERCENT OF CORRECT

									R
1	89	96	97		89	93	93	97	93	8
2	115	114	107		101	107	105	107	108	14
3	102	76	96		89	99	102	107	96	31
4	95	106	101		114	102	100	111	104	17
5	94	97	117		99	128	122	115	108	34
6	88	90	98		69	108	108	96	94	39
7	95	95	119		119	112	119	119	109	24
8	98	92	91		96	91	86	93	93	12
9	96	104	103		90	99	96	96	97	14
10	97	123	119		117	124	131	130	121	34
$\bar{X}$	97	99	105		98	109	106	107	102 = $\bar{X}\bar{t}$ $\bar{R}f$ = 23
R	27	47	28		50	37	45	37	39 = $\bar{R}m$

ANALYSIS BY MEN

$\bar{R}m = 39$ $\bar{\bar{X}}m = 102$

$UCL\bar{X}_m = \bar{\bar{X}}_m + A_2\bar{R}m = 102 + .308 \times 39 = 114$

$LCL\bar{X}m = \bar{\bar{X}}m - A_2\bar{R}m = = 90$

$UCLRm = D_4\bar{R}m = 1.777 \times 39 = 69$

$LCLRm = D_3\bar{R}m = .223 \times 39 = 9$

ANALYSIS BY PACE

$\bar{R}f = 23$ $\bar{\bar{X}}f = 102$ $SD = \frac{\bar{R}f}{d_2} = \frac{23}{2.849} = 8.5$

$UCL\bar{X}f = \bar{\bar{X}}f + A_2\bar{R}f = 102 + .373 \times 23 = 112$

$LCL\bar{X}f = \bar{\bar{X}}f - A_2\bar{R}f = = 92$

$UCLRf = D_4\bar{R}f = 1.924 \times 23 = 44$

$LCLRf = D_3\bar{R}f = .076 \times 23 = 2$

Figure 4.

M. E. Mundel 570115-5

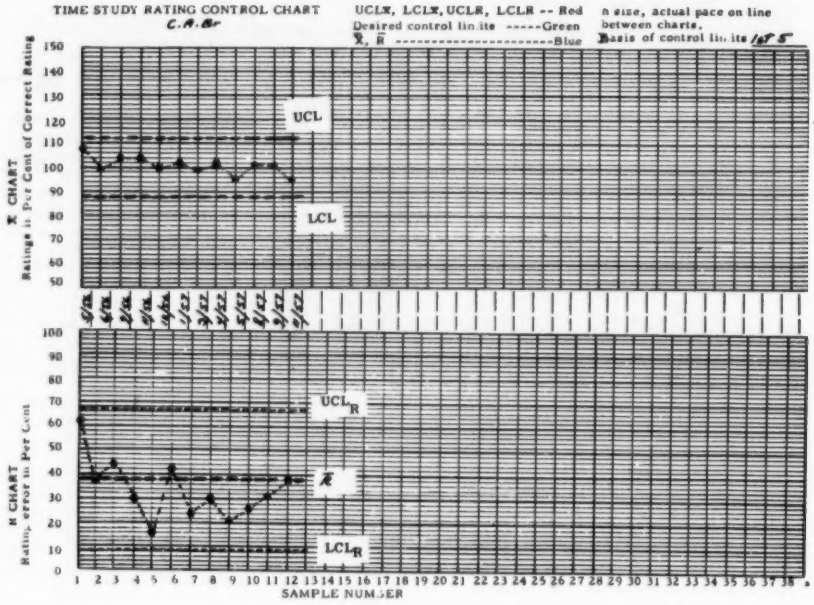


Figure 3. Control chart for one time study man from May 1956 through November 1957.

Subsequently, ten films, one film from each of ten series of films may be presented for rating. With ten new films each month it is deemed unlikely that even time study men can spot which films were selected for this mixture. Results of such a session are shown in Figure 4.

Of first importance is the value of K. The A column in this case consists of the appropriate KA values from the first rating session, hence K now represents how well the concept of the unit of measure, previously used in the ten rating sessions, was retained. The individual R and $\overline{X}$ for each man and for each film indicates, as any sample R or $\overline{X}$, either a control condition (only chance causes within the control limits) or an assignable cause, for analysis. In the example shown here, film number ten was greatly over-rated. Subsequently the group would review this film and seek the cause of their error. In such a manner, they may learn to improve their judgement process. In the plant from which these data were taken these techniques have assisted in reducing the error in ratings to about 20% of what they originally were, as well as eliminating bias differences among time study men and speeds of work.

This has been far from a complete description of all uses of quality control techniques in time study activity. For non-repetitive activities or tasks, true sampling techniques are used to assist in determining both allowances and standards. The methods employed require a considerable description which would cause this paper to be undesirably lengthy.

Suffice to say that those features which have made quality control techniques more suitable for inspection and control work than classical statistical procedures, likewise ideally suit them to time study activity. Certainly this is a basis for coordinate effort between the quality-control group and the time study group.

1. Mundel, M. E., "Motion and Time Study," Prentice-Hall, N. Y., 2nd Ed., 1955, pps. 306-308.

2. Mundel, M. E., Op. Cit., page 316.

3. This is only a very brief statement of the basis of excluding values. See Mundel, M. E., Op. Cit., pps. 347-349.

4. A method of rating called "Objective rating" was employed. For a description of this method see, Mundel, M. E., "Improving Time Study Techniques," Proceedings 1955 Time and Motion Study Clinic, Industrial Management Society, Chicago, Ill., pps. 1-8. With conventional rating techniques, vastly different results are obtained with many "out of control" values.

TRANSFORMATIONS OF THE INDEPENDENT VARIABLE

G. E. P. Box
Princeton University

SUMMARY

In problems of regression analysis and also in response surface work it is frequently the case that more adequate representation of a function, relating a dependent variable η to the levels of a number of independent variables x_1, x_2, etc. by means of a simple model such as that provided by a linear or quadratic equation, would be provided in terms of transformed independent variables. Thus a linear model in $\log x_1$, $1/x_2$, etc. might be adequate in many circumstances to represent a slightly curved but monotonic surface. A method is described whereby the best transformations can be obtained from the data themselves, and confidence limits showing the range of transformations uncontradicted by the data can be constructed.

SENSORY TESTING

George E. Ferris
Staff Statistical Specialist
Post Cereals Division
General Foods Corporation

I. INTRODUCTION

The subject of Sensory Testing interpreted in the broad sense covers much more ground than is appropriate for discussion before a Society for Quality Control. Therefore, after outlining the various situations in which the need for sensory testing arises, attention will be confined to those where sensory testing is used as a tool for controlling quality.

There are essentially five different types of problems which can employ sensory testing. These are:

1. New product development.
2. Cost reduction problems.
3. Quality improvement problems.
4. Evaluation of popularity against competition.
5. Quality control problems.

The first four of these are discussed in references (1) and (2) and a summary of the literature available on these topics is given there. The fifth problem can be further broken down as follows:

5. (a) Studying the effect of variation in the source, nature, or proportion of the ingredients.

   (b) Storage studies, shelf-life studies.

   (c) Routine, day to day quality control, including inspection of both incoming raw materials and outgoing finished products.

These will be the problems discussed. Within the context of 5 (a), (b), and (c), two different practical situations may arise:

   (i) The purpose may be to investigate whether an appreciable difference exists between a variant and a standard.

   (ii) The existence of an appreciable difference between the variant and the standard may be admitted, but it is hoped to show that despite the difference there is no difference in consumer preference, or is in favor of the variant.

It is evident that in situation (i) the solution lies in the use of an expert or analytic sensory panel, administered internally by the research laboratory concerned, whereas, in situation (ii) a consumer preference (sensory) panel has to be used, and this is generally done by the research laboratory through some outside agency. Thus, in case (ii) even if the expert panel finds an appreciable difference, a consumer test

will still be run, or the expert panel stage is simply omitted.

II. ADMINISTRATION AND ORGANIZATION

1. The role of the statistician

When we are consulted, our first job is the identification of the problem. The diagnosis of a quality control problem into 5 (a), (b), or (c), type (i) or (ii) must be done after a careful cross-questioning of the research workers who regrettably are not always clear in their own minds about the nature of their problem. An incorrect diagnosis may result in excellent medicine for the wrong disease. At this stage also, sufficient information must be elicited by us from the research worker (or management) so that we may assess the consequences of the Type I and Type II errors associated with every statistical quality control design (also known as producer's risk, consumer's risk, see reference (3) for definition); e.g., (i) what will be the consequences if the expert panel incorrectly allows a certain percentage of product outside specifications to be marketed; (ii) what loss results if the panel incorrectly rejects the variant as appreciably different from the standart when, in fact, it is not?

Armed with this knowledge, we can prescribe the appropriate type of test of the required magnitude. We have available a long list of techniques for both the expert and the consumer panels. A difficulty we often face at this stage is to convince our client (or patient) that unless he runs the prescribed size of test (number of judgments) he may just as well not run the test at all, because a smaller test simply will not give him the information he seeks. In the light of the researcher's lack of funds, the Type I and Type II errors often have to be revised.

The actual testing is then carried out by the appropriate panel while the statistician closely supervises it to ensure no biases creep into the procedure. Possible biases will be mentioned as each technique is discussed.

Finally, after data are assembled, we analyze and report the results. We will try to foresee every possible way in which management or the researcher may misinterpret or unduly generalize our conclusions and protect ourselves against it in our report.

2. Facilities inside the laboratory

(a) Physical setup

The physical facilities that should be available for expert, analytic, or booth panel work include booths for individual judges (to ensure independence), noise and odor-free surroundings, lighting facilities such as to enable colors to be masked, and dishes and utensils as required by the product to be tested, and ample preparation area. There are numerous references in literature, see references (2) and (4), indicating the importance of each and every such detail on the efficient operation of the panel. The organizer of the panel

should have the authority to penalize the judges not keeping their appointments to taste without excuses accepted as valid, see reference (5).

(b) Basic information on members of staff

In order to help the person responsible for the organization of the expert booth panel, he should have certain information on the staff of the laboratory or plant who are available to act as judges on the panel. This information might be on IBM cards for each eligible judge. In addition to age, sex, ethnic background (especially circumstances during the habit-forming years of the early teens), the information should include medical data on certain diseases which disqualify anyone from acting on an expert booth panel, see reference (3). To this, the organizer adds results of performance on threshold-recognition and near-threshold differentiation-tests with the four primary tastes, the nine primary odors and combinations of them. On this card too, can be kept the running records of the judge's performance on various products once he becomes an active member of the panel.

(c) Panel selection, education, and maintenance

From the basic information on individuals, certain members of the staff can be disqualified from acting as judges by reason of health or inability to distinguish fundamental differences. Other criteria become operative once the judge joins a panel. His performance is watched by rating duplicate samples (if his two ratings disagree too badly too often, he is deposed), on ranking test, matching tests, pairing tests, and multiple choice tests described below. If he departs too far from the panel average too often, or if he displays overconservativeness or extremism on rating scales, he is shown the error of his ways and if he does not improve, he is sadly bidden farewell. Control chart methods are practiced in some laboratories on the mean and variance of individual judges, similar to quality control methods. To replace judges who fall by the wayside, there must be available a pool of new judges; these can be members of staff who have been called on when regular members have been absent. Their places as "emergencies" are taken by new candidates. Often an "emergency" does so poorly so often that he is found out before becoming a member of the regular panel, but before so disqualifying anyone he should be tested out in a reasonable number of tests.

(d) Expert (booth) panels

These panels consist of those members of the entire staff who have been chosen for their ability to detect and identify the presence of primary flavors in near-threshold concentrations and their ability to discriminate between small differences in concentration, together with ability when tested with a series of variants of the product being studied. The panel is used as a piece of equipment for measuring differences among samples submitted to it, for fundamentally two purposes:

(i) To show that some of the samples differ significantly.

(ii) To show that samples submitted do not differ appreciably.

(e) Other laboratory panels

Informal bench-top panels are often used as precursors to formal booth panels, but not usually on quality control projects. Profile panels comprise members of staff selected on their performance in booth panels and on the basis of a good taste vocabulary. They are often trained by outside experts (flavor houses) in the art of taste-testing. Mostly they are used in non-quality-control projects, but they are often useful when the booth panel finds a difference (e.g., in the case of an unsuccessful 5 (a) attempt to duplicate a standard, using new ingredients) by explaining or describing the nature of the difference. Differences in preference from a consumer test can similarly be explained.

3. Facilities outside the laboratory

Whereas, problems 1 through 4 listed in the Introduction may require a great variety of consumer panels, all the way from "depth interviewing" to nationwide test marketing, these will be found described in reference (1); for quality control purposes, the conventional type of consumer taste test is sufficient. For quality control purposes interest lies in the product, unconfounded by the influences of labelling, packaging, etc.

The principal objective is to estimate the percentage of the population who have no real difference in preference between the standard and the variant and the proportions of consumers who prefer each. This can be done by getting consumer judgments in places like department stores, theaters, ballgames, schools, prisons, factories, and offices. The location chosen and method of questioning used, are varied to fit the purposes of the test as regards the consumer population required to be sampled and the question to be settled. The questions asked usually pertain to preference or acceptability. The underlying design is generally simple or multiple paired comparison, but single evaluations are sometimes done this way too (see below). The technique known as the repeated pair or 2-visit procedure can be used to advantage, see reference (6).

A consumer panel must be selected bearing in mind the following considerations:

(a) It must be large, heterogeneous, representative of the population consuming, or likely to consume, that product which is being investigated (obviously not the same for champagne as for bread, nor the same for chili-con-carne as caviar).

(b) Its members must not be trained or over-instructed.

(c) An estimation of the number of guessers must be possible,

because repeated evidence is available that panel-members tend to vote for preferences "just to oblige."

In particular, no attempt should be made to exclude non-discriminators on grounds either genetic (taste-blind), medical, nor any similar reason if in the first place these members were selected by a random or other statistical process, as (a) would be violated otherwise. In view of (b), the use of "standing home-use-panels" is much to be discouraged especially for quality control purposes as members may get over-familiar with product. Finally, (c) makes necessary the repeated pair procedure, see reference (6).

III. STATISTICAL TECHNIQUES AND CONSIDERATIONS

1. Multiple choice tests

A very convenient tool for us is the class of designs known as multiple choice tests. Chapter 1 of Fisher's, "Design of Experiments," see reference (7), has the illustration of the lady who claimed she could detect whether tea was made with milk poured in before or after the tea. Fisher proposes several designs to confirm or reject the lady's claim, all of which are of the multiple choice type.

Such tests consist of asking judges to identify or group samples into classes without asking them for preference, e.g., judges may be told: "Here are 6 samples; 3 are alike of one kind, 3 are alike of another kind; please do your best to sort the 6 samples into 2 groups of 3," or "here are 5 samples, 4 are alike and one is different; please identify the odd one." In each case, the probability of grouping or identifying correctly by pure guess is readily calculated (1/10 for the first example, 1/5 for the second). If the difference among samples is real, or if the judges have an ability beyond a mere guess, then the proportion of correct answers will be significantly greater than the chance probability. A notable example of this type of test is the test variously known as the triangle, odd-sample, triad, delta, or trio test. This test is enjoying great popularity and is often used even when it should not be. It is one of the simplest multiple choice tests possible and consists of presenting sets of three samples to a panel of N judges ($N = 6n$, $n = 1, 2, 3, \ldots$). Each judge gets a set of the form AAB, i.e., 2 alike and one odd. They are told that 2 of their samples are identical, one different, and they are asked to identify the odd one. Of the 6n judges, n get the samples in the order AAB, n in the order ABA, BAA, BBA, BAB, ABB respectively, to balance possible order and residual effects. If the number of correct identifications is significantly greater than the 2n expected by pure chance, it is concluded that samples A and B are different.

This technique, with all its variations, solves most quality control problems other than storage and shelf-life problems. We can vary the size of the expert panel, the number of judgments per panel member, the basic design of the test -- triangle, pentagonal, double-tetrad, etc. -- and hence, the probability of guess-identification. In addition, we can work with a pre-fixed total number of judgments, or a pure sequential plan, or getting groups of judgments

with truncation at some pre-fixed total number. A very popular design is sets of 6 judgments of the triangle type with truncation at e.g., 60 judgments.

There often is difficulty in explaining to researchers the need for a large number of judgments when the problem is of 5 (a) or 5 (c) type (ii), i.e., when the purpose is to show no significant difference. This sequential procedure can be diagrammed and together with the probabilities of incorrect conclusions of each type helps convince him of the need for a large number of judgments.

The choice is, of course, governed by Type I and Type II error considerations. It is an unfortunate property of multiple-choice tests that their Type II error is very difficult to interpret, see reference (8). However, once an agreement is reached on Type I and Type II errors between the parties concerned (supplier and the laboratory on incoming material, research and marketing on outgoing material) the design to be adopted by the statistician is clear cut.

There are times when multiple choice tests are too uneconomical from the point of view of number of judgments required. A numerical measure may be desired of the magnitude of the difference of a particular characteristic. Below are some alternative techniques.

2. Rating scales

Multiple choice tests avoid all difficulty that arises from trying to get a numerical measure of the intensity of a characteristic, or of the difference between samples.

It is sometimes imperative to rate in booth panel work the flavor-intensity, say, of a number of samples. To this end rating scales with varying numbers of steps are used. Judges rate a number of samples on a number of characteristics, each on such a scale. In conjunction with such scales the actual design may be any one of the classical designs, randomized block, latin square, incomplete blocks, see references (9) and (10). Which of these is used depends on the total number of samples to be intercompared and the panel size required from Type I and Type II error considerations. The analysis is performed on totals over judges' ratings. The panel as a whole is regarded as a piece of equipment; the individual judges' ratings or judge x sample interactions are of no interest whatever, except when the judge is the incomplete block of the incomplete block design, when judge effects are corrected for.

The rating scale can be used in a somewhat different way; e.g., when a number of samples are being tested for overall difference from a control sample. In such cases, one end of the scale reads "no difference," the other "completely different," and the number of steps may still vary. It is usual to throw in a "blind" control or to replicate one of the variants as a check on panel performance. The scale used in conjunction with consumer tests is of a different kind.

3. Matching tests

A very useful device in booth panel work is the method of matching, illustrated by the following situation:

Suppose the five known solutions x_1, x_2, x_3, x_4, and x_5 are presented to a panel with a sixth sample z, which is thought to have concentration in the range x_1 to x_5. Each judge is asked to match z with one of the x's. This enables an estimate of the concentration of z to be made. Should z be known, the results enable judges to be evaluated.

This is another method where the difficulties of using a scale are avoided, and still very useful in quality control work.

4. Ranking tests

Still another method for booth panel work which avoids scales is illustrated by the following problem:

Suppose there are five sugar solutions of concentrations x_1, x_2, x_3, x_4, and x_5 respectively, which are known to form a geometric progression. The five properly coded samples are presented to a panel of n judges, each of whom ranks them in what he thinks is the increasing order of concentration. Such rankings enable judges to be rated. Alternatively, suppose the concentrations of x_1, x_2, x_4, and x_5 are known, and the fifth sample $y = x$ has an unknown intensity believed to be between x_2 and x_4. By getting the judges to rank this unknown in with the known concentrations, an estimate of its concentration can be obtained from the average ranks.

5. Pairing tests

A very similar procedure to 3 and 4 immediately above can also be used. Suppose we have the same 5 solutions as in 3 and 4, together with a solution w of unknown concentration available. It is necessary to estimate the concentration of w. The panel is presented the pairs x_1 and w, x_2 and w, x_3 and w, x_4 and w, x_5 and w, in turn properly coded. The pairs are presented in a random order as are the samples within a pair. At each session, judges are asked to rate which sample of the pair is higher. This leads to an estimation of w. Judges can also be so evaluated.

The techniques below are principally for consumer tests which can occur as part of quality control programs.

6. Single (or monadic) evaluation

In consumer tests, it is not always advisable to evaluate products several at a time. When a product is so unusual that there is no reference point to compare it against, then single evaluation is the most sensible procedure. The samples are sent out to members of the consumer panel and they rate the product on some scale as the one below, either for overall preference or for individual properties (like sweetness, fullness, ease in preparation, etc.). The ratings are not taken absolutely at face value, but provide a guide as to direction of advantageous modifications, and often put the

researcher in a position to assess several experimental variants relatively to one another.

Typical scale for single evaluation

Verbal description	Quantitative measure
I like this sample extremely well	+3
I like this sample moderately well	+2
I like this sample slightly	+1
I neither like nor dislike this sample	0
I dislike this sample slightly	-1
I dislike this sample quite a bit	-2
I dislike this sample extremely	-3

7. Paired comparisons

The method of paired comparisons is an especially useful tool in consumer taste-testing. It consists of presenting to the consumer panel the pair of samples under investigation and asking them to indicate preference for one or the other, or to say "no preference." Occasionally, those who indicate a preference are further asked to indicate the degree of their preference, which can then be quantized as shown below:

Scale to be used in conjunction with paired comparisons

Verbal description	Quantized measure
I like A very much better than B	+3
I like A quite a bit better than B	+2
I like A slightly better than B	+1
I like A and B equally well (or badly)	0
I like B slightly better than A	-1
I like B quite a bit better than A	-2
I like B very much better than A	-3

Care is taken to balance order of presentation effects and code biases. The results are then analyzed, see reference (11), and it is established whether there is a significant preference for A, or B, or no significant preference. The procedure, except for the length of questionnaire, is the same for problems of type 1 through 5 (see Introduction) both in 7 above and 8 below.

8. Multiple paired comparisons

When some number (say t) of samples are submitted for consumer preference evaluation ($t > 2$), the most efficient method is that of multiple paired comparisons. This consists of evaluating each of the $\frac{t(t-1)}{2}$ paired comparisons by the method in 7 immediately above. This is done even when direct interest may only be in some of the paired comparisons, as the indirect information from the others contributes towards the samples of greatest interest, and in this sense a balanced design, i.e., and equal number of judgments on all

paired comparisons is the most efficient. The votes are then tabulated and analyzed by one of two standard methods. In the Bradley-Terry procedure, see references (12) and (13), the preferences are analyzed as ranks ("no preference" = tied rank); if degree of preference was recorded, we may use the Scheffe method, see reference (11), which gives a little more information, though a Bradley-Terry analysis can still be done. The Scheffé procedure treats the actual scores from a scale similar to one shown in 7. immediately above.

9. Modified duo-trio tests

This is the first of two methods described here which in consumer tests enables the investigator to assess the proportions of consumers preferring A and B and the proportion having no real preference, by not only asking them but by checking their statement by their performance. The modified duo-trio test, see reference (14), on a given pair of Products A and B consists of presenting to all N judges of a consumer panel one of the following sets of samples:

To $\frac{N}{4}$ judges: A and B, then A

To $\frac{N}{4}$ judges: B and A, then A

To $\frac{N}{4}$ judges: A and B, then B

To $\frac{N}{4}$ judges: B and A, then B

After tasting the first pair, the judge is asked to indicate a preference (or no preference). He is then given a third sample and told: "This third sample is identical to one of the two you have just tasted: please indicate which one it matches." The preferences of those who match correctly are tabulated separately from those that do not. The preferences from the former tabulation are clearly more reliable and can be further adjusted by assuming that some of those who did match correctly, did so by pure chance. It is important to note that all judges gave their preference before being asked to match -- this can be psychologically important.

10. Repeated pair tests

A second method, see reference (10), which provides reliable and revealing information consists of getting the panel of N judges (or some part of it) to indicate preference on the same pair of samples twice (without their knowing it). From the number of reversals of preference, and the distribution of the other votes, we can get a better mathematical estimate than from the single paired comparison of that part of the consumer population which has no consistent preference. It also enables us to distinguish between the case where there is no significant preference for either A or B because the population is equally and perhaps sharply divided, and the case where there is no significant difference in preference because the entire population is indifferent to the magnitude of the

difference between the samples or even unable to distinguish between the samples.

Putting it another way, the estimates of the proportions of preference for each product and no preference are made practically independent of the method of wording the questionnaire or interview, as the case may be: This, of course, is a considerable improvement over other procedures.

11. Analysis of variance in storage problems

Storage studies and shelf-life studies may require a slightly different approach to those discussed so far. In these situations there is usually a reference sample, being either fresh material, or material produced at the same time as the other stored samples and kept extra well refrigerated. The other samples are material stored at various temperatures for varying periods. Samples are evaluated by one of two overall plans.

(a) One may arrange for all samples to come out of storage at the same time. This can be done by producing all the material at a given time, keeping samples at a low temperature, and gradually transferring them into higher temperature storage at the appropriate times, so that at a calculated time all samples can be withdrawn and compared by one of the procedures 1 to 10 above.

(b) One may put all samples into storage at the selected temperatures and make withdrawals and evaluations (on a scale against a standard) at various times. For each batch of material, a regression coefficient with its sampling error can be determined (regression of rating on time of storage -- transformed, if necessary, to be linear). An analysis of variance may then be performed on the regression coefficients to determine whether time and/or temperature significantly affect ratings.

12. Correlation with objective measurements

It is sometimes convenient to establish the correlation between the acceptability of the product to the consumer and a physical, objectively measurable, property thereof. For example, if the acceptable range for a given cereal is shown to be 1.2% to 2.7% moisture, with other properties fixed, one may exercise quality control on moisture rather than on a sensory test. It is more usual to replace a sensory test not by one, but by a combination of objective measurements.

13. Discussion

It is not enough for us to be aware of the existence of the above techniques. We have to know the right occasion to employ each and how to handle certain non-statistical problems or semi-statistical problems connected with each.

Some of the decisions we have to make at the booth panel stage

are these:

(a) To choose between a multiple choice test or a rating test (less often matching, ranking, or pairing).

(b) How many samples can be tasted by each judge at one session without undue fatigue?

(c) What, if any, control samples should be included in the test?

(d) Should central or end standards of reference be included?

(e) How many blind controls or duplicated variants should be included to check judge performance?

(f) How should the questionnaire be worded?

(g) How many steps should each rating scale have?

(h) Is it best to rate individual characteristics and perhaps combine weighted ratings at the finish, or should a single overall rating be obtained?

(i) Is the material uniform and representative?

(j) How much information (background) should be given to panel members?

(k) Should descriptive adjectives be used on the scales?

(l) Should each judge's scores be analyzed separately, or should all scores be combined -- straight or weighted?

(m) Should the scores or ratings be replaced by ranks or should they or the time-scale be transformed? (To make the scale homo-scedastic, or the scores normally distributed, or regressions linear).

(n) Are the ratings additive? Multiplicative? Weightable?

(o) What is the best model for the analysis? For example, if the samples form a factorial design (on processing factors say) is the split plot model applicable, or some other?

The decisions to be made in connection with consumer tests, if run for quality control purposes, are comparatively easy.

(i) The consumer population aimed at must be defined.

(ii) The method of obtaining a representative sample (the frame) must be chosen, i.e., the type of consumer test most applicable must be selected.

(iii) A short, bias-free questionnaire or method of interviewing must be made.

REFERENCES

(1) Ferris, G. E., Taste panels. BU-61-M, March 1, 1956. A publication of Biometrics Unit, Cornell University.

(2) Dawson, Elsie H., and Harris, Betsy L., Sensory methods for measuring differences in food quality. Bureau of Human Nutrition and Home Economics, USDA, Ag. Inf. Bull., 34, 1951.

(3) Dixon, W. J., and Massey, F. J., Introduction to Statistical Analysis. McGraw-Hill Book Company, New York, 1951.

(4) Peryam, D. R. et alii. Food acceptance testing methodology. National Academy of Sciences, National Research Council, Symposium, October, 1954.

(5) Hopkins, J. W., Instructions as to procedure in laboratory tasting. Can. J. Res. XXIV and XXV, No. 1, January, 1946.

(6) Ferris, G. E., The k-visit method of consumer testing. Biometrics 14, No. 1, March, 1958.

(7) Fisher, R. A., The design of experiments. Oliver and Boyd, Edinburgh, 4th edn., 1947.

(8) Ferris, G. E., Taste-testing quality control. Rutgers Meeting of the American Society for Quality Control, September 5, 1957.

(9) Kempthorne, O., The design and analysis of experiments. Wiley and Son, New York, 1952.

(10) Ferris, G. E., A modified latin square design for taste testing. Food Research 22, No. 3, 1957.

(11) Scheffé, H., An analysis of variance for paired comparisons. Journal of the American Statistical Association 47, 1952.

(12) Bradley, R. A., and Terry, M. E., The rank analysis of incomplete block designs. Biometrika 39, 1952.

(13) Dykstra, O. Jr., A note on the rank analysis of incomplete block designs. Biometrics 12, 1956.

(14) Bradley, R. A., The modified duo-trio method. Unpublished correspondence, 1956.

SOME PRACTICAL APPLICATIONS OF SAMPLING

George Vlases, Jr.
U. S. Treasury Department, Bureau of Customs

The handling of commercial importations of merchandise for duty purposes is one of the most important functions of the United States Treasury Department's Bureau of Customs. The number of such transactions has more than doubled in the last ten years. The increased workload has been handled without any increase in personnel by the use of a number of improved management techniques. One of these is the application of probability sampling methods to the testing, sampling, and weighing of imported merchandise.

Sampling, weighing, and testing are carried out in the various ports by the customs appraisers, collectors, and chief chemists, respectively. The supervision of this work is under the Commissioner of Customs in Washington, D. C., and is delegated to several of eight divisions. One of these, the Division of Technical Services, has as one of its tasks the supervision of testing. In addition, it has the responsibility of furnishing information on probability sampling to the Bureau as a whole.

We have made extensive use of analytical statistics in testing. The kinds of commodities required to be tested are extremely numerous, and our methods cover nearly all fields of chemistry. We have found a knowledge of the standard deviations for single tests to be useful in our work; in making decisions at a class dividing line, and in evaluating the acceptability for customs purposes of tests made by commercial laboratories.

Customs uses a large number of analytical methods and has used a number of procedures to economically accumulate values of the standard deviations of tests. One of these is our interlaboratory test program. About eight of the most important customs analytical chemical methods are tested each year for precision and, in some cases, for accuracy. Nine laboratories usually participate. The design of the experiment and interpretation of results are based on statistical methods. Another source of estimates of these standard deviations is our methods development program. Such analytical chemical methods as cannot be tested by the interlaboratory test program using regular random routine conditions are tried out by a single experienced chemist under the best controlled conditions. We have found, as a rough generalization, that standard deviations obtained by the first method are about twice those obtained by the second. Because of the large number of chemical test methods we have to use, however, we have found it most convenient to obtain estimates of the standard deviations of tests by running a replicate on a given sample at another time. It is preferable for our purposes to do this on another day, under routine conditions, and without the knowledge of the chemist. If it is not possible to economically conceal the identity of the replicates, objectivity can be obtained with training. The standard deviation of individual tests, s_1, where two replicates are used, is given by:

$$s_1 = \sqrt{\frac{\Sigma(X_1 - X_2)^2}{2m}} \quad \text{where}$$

X_1 and X_2 are the values of tests on replicate pairs, and m is the number of such pairs. This formula has the advantage of being reasonably effective over a considerable range of values of the average $\bar{X}$. Good results for s_i have been obtained where metal contents in ores varied from 50 to 60 percent.

One of the principal uses of the data on the standard deviation of tests has been to serve as the basis for decision when a sharp tariff class dividing line is involved. There are many provisions in the Tariff Act, involving such commodities as alloys, alcoholic beverages, vegetable residues, drugs, ores, and others, where there is an abrupt change in duty rate at a specific value of some measurable characteristic. For one particular ore, for example, there is provision for a fourfold change in duty when the content of active component exceeds 97.00 percent. From a knowledge of the standard deviation of tests, the number of replicates necessary to establish, with specified reasonable assurance, whether or not the component in the sample exceeds 97.00 percent can be readily calculated.

Another use we have made of the values of the standard deviations of tests has been to judge the acceptability of commercial settlement tests for duty purposes. For many ores and alloys, assays are made by the buyer, the seller, a commercial referee laboratory, and customs. The settlement figure (for payment) is often arrived at from a consideration of the first three tests. For certain commodities we have, over the years, accumulated considerable data on the standard deviations of tests by others and by customs. These data enable the chief chemists of customs to make recommendations as to the reasonableness of observed differences between settlement and customs tests. Such recommendations may lead to the acceptance of settlement tests for duty purposes with resulting large savings in paperwork.

Probability methods have been used extensively by customs in sampling merchandise for test. The principal benefits have been the determination of the most economical sample sizes for a predetermined precision. Considerable success has been achieved, particularly with natural and synthetic textile fibers.

Probability sampling has been used by customs to help reduce the volume of comptroller's verifications of customs transactions, in the inspection of parcel post packages, to assist in the prediction of the volume of future business, and to help evaluate the difference between certain customs procedures. In some instances, a combination of random sampling (to estimate possibilities or levels, as required) with judgement sampling (to obtain most effective performance) has been found useful.

One of the most fruitful applications of sampling to customs work has been the sample-weighing of large importations of packaged cargo. This was suggested in the appendix to a management report by Dr. W. Edwards Deming in 1948.

The Customs Service, originally responsible for standards of weights and measures, had, over a period of about one hundred years, developed weighing procedures of which it was justifiably proud. These called for accurate 100 percent weighing, and the proposal for weighing by sample

was an almost complete departure from proved and established procedure. Despite this, cooperation of personnel was obtained by first selling the new proposals to a few key supervisors in charge of weighing operations. The new procedures were also attractive because they gave promise of releasing personnel for other necessary work. Importers were sold on the new weighing proposals since it appeared their merchandise would move faster through customs. Because of this, they readily made certain necessary weight documents available to customs.

The most important feature of the sample-weighing methods is the use of the conversion factor technique. Much of the merchandise imported has been weighed abroad, and the use of a conversion factor on the foreign weights is necessary to evaluate unidirectional changes in weight.

If an importation has been previously weighed, and the weights of the individual packages or containers comprising the lot are available, we can use sample-weighing to arrive at the current weight, which may be substantially different. By taking a random sample of the packages, we find that there are certain differences in the corresponding pairs of individual customs and importers weights. These may be due to changes in moisture content, differences in weighing equipment or techniques, and other causes. These differences, customs minus importer (or invoice) weights for the packages taken into the sample, C-I, can also be evaluated by the $\Sigma C / \Sigma I$ which is equal to f, the conversion factor or ratio estimate. The sum of all the importers (or invoice) weights in the lot times the conversion factor gives the current weight for customs purposes.

Assume a sample of N differences, C-I = d, where C is the customs weight of a container and I is the importers (or invoice) weight of the same container. Let this sample of N items consist of y subgroups of size x. Then, if R_x is the range of differences in any subgroup

$\Sigma R_x/y = \bar{R}_x$, the average of the ranges of differences.

Further, $\bar{R}_x/d_2 = s_d$, the standard deviation of the differences, where d_2 is a constant depending on subgroup size (1).

The conversion factor or average ratio from the sample,

$$f = \Sigma C/\Sigma I = \bar{C}N / \bar{I}N = \bar{C} / \bar{I} = (\bar{d} + \bar{I})/ \bar{I} = \bar{d} / \bar{I} + 1,$$

where the bar, $\bar{}$, indicates the average value.

The standard deviation of the ratio C/I, s_f, is estimated by

$s_f = s_d / \bar{I}$, approximately (where the value of f is near one)

$s_f = \bar{R}_x/d_2\bar{I}$

The coefficient of variation, in percent

$$V_f,\% = 100\ s_f/f = 100\ \bar{R}_x\bar{I}/d_2\bar{I}\ \bar{C} = 100\ \bar{R}_x/d_2\bar{C}$$

The standard error of the conversion factor, in percent (E,%), is obtained from

$E,\% = tV_f/\sqrt{N} = 100\ t\ \bar{R}_x/\sqrt{N}\ d_2\bar{C}$ where

t is a probability factor taken equal to 3.

$$\bar{R}_x = \frac{d_2\bar{C}\ E\sqrt{N}}{100\ t}$$

The use of the above formula for the average of the ranges of differences will be illustrated by the weighing procedure for cigarette leaf tobacco, an attachment to Bureau of Customs Circular Letter No. 2655, May 5, 1949. This commodity comes from the Near East and adjacent territory. Turkish tobacco comes in bales averaging from 90 to 135 pounds for different lots. Greek tobacco bales average around 50 to 70 pounds and Latakia tobacco averages 250 to 330 pounds. For customs weights alone, within any one importation of these tobaccos, the coefficients of variations of the sample-weights were too high, making the use of an average customs bale weight, based on a weighed sample, unsatisfactory for the precision desired. We found, however, that where we took a sample of corresponding pairs of customs weights and invoice (importers or foreign) weights, the coefficient of variation of the conversion factor was sufficiently low and stable to give the desired precision. The table below, Control Table for Use in Determining the Total Number of Bales of Cigarette Tobacco to be Weighed by Customs, was calculated using this information and the formula for $\bar{R}_x$. The latter becomes $\bar{R}_5 = 0.001938\ \bar{C}\ \sqrt{N}$ where $x = 5$, $d_2 = 2.326$, $t = 3$, and the desired precision, $E = \pm 0.25\%$. While a sample of $N = 100$ ordinarily suffices, provision is made for additional sampling, where needed, to approximate the desired precision.

A specific example for cigarette tobacco is given on a reproduction (left hand column for Mark & Bale No. omitted) of customs Form 6077, below. The lot has been assigned the designation Warehouse Entry 5072. Only the last 20 of the 100 individual bales comprising the sample are shown, but this is adequate to illustrate the procedure. The customs weights are recorded on the spot. For example, bale No. 6724 showed a customs weight of 138.5 pounds. Its invoice or foreign weight, converted from kilograms, is 138.3 pounds. In getting the range of differences for any subgroup, we obtain the spread between the largest negative and positive numbers. Differences, C-I, may be positive or negative. Ranges of differences are always taken as positive. The method of calculating the average range, the gross, and net are evident from the form.

The value of $\bar{R}_5$ actually obtained, 0.89, tells the weigher that the sample of 100 is satisfactory since, from the control table below, for $N = 100$ and $\bar{C} = 130$ (approximately) the limiting value of $\bar{R}_5$ for $E,\% = \pm 0.25$ is 2.5. Had we obtained an actual observed value from the sample of $\bar{R}_5$ larger than 2.5, say 3.9, it can be seen from the table that N must be 250 to approximate the desired precision. In such a case, an additional 150 bales would be weighed.

The weighing returns are reviewed for statistical adequacy and possible gross error. The former is carried out by the usual methods. Gross error can usually be detected by a study of the total invoice weights, the total number of bales, the average sample-weights, customs and invoice, and the standard deviation of the customs weights in the sample.

The above procedure has been applied to the weighing of a number of bulk commodities. It is evident from the basic equation for $\bar{R}_X$, above, that control tables to fit the required circumstances are readily set up. Initial sample sizes as low as 20, ten subgroups of 2, have been found satisfactory. The selection of precision levels (values of E,%) is a matter of policy to be guided by considerations such as the costs of weighings and the magnitude of possible deviations in duties charged.

CONTROL TABLE FOR USE IN DETERMINING
THE TOTAL NUMBER OF BALES OF CIGARETTE
TOBACCO TO BE WEIGHED BY CUSTOMS

$\bar{R}_5 = 0.001938 \times \bar{C} \times \sqrt{N}$, for E = ±0.25%

($\bar{C}$) Customs Average Bale Weight Pounds	Number of Bales to be Weighed (N)									
	100	150	200	250	300	350	400	450	500	550
	Average Range of Differences in Groups of Five ($\bar{R}_5$)									
30	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.2	1.3	1.4
40	0.8	0.9	1.1	1.2	1.3	1.5	1.6	1.6	1.7	1.8
50	1.0	1.2	1.4	1.5	1.7	1.8	1.9	2.1	2.2	2.3
60	1.2	1.4	1.6	1.8	2.0	2.2	2.3	2.5	2.6	2.7
70	1.4	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.0	3.2
80	1.6	1.9	2.2	2.5	2.7	2.9	3.1	3.3	3.5	3.6
90	1.7	2.1	2.5	2.8	3.0	3.3	3.5	3.7	3.9	4.1
100	1.9	2.4	2.7	3.1	3.4	3.6	3.9	4.1	4.3	4.5
110	2.1	2.6	3.0	3.4	3.7	4.0	4.3	4.5	4.8	5.0
120	2.3	2.8	3.3	3.7	4.0	4.4	4.7	4.9	5.2	5.5
130	2.5	3.1	3.6	4.0	4.4	4.7	5.0	5.3	5.6	5.9
140	2.7	3.3	3.8	4.3	4.7	5.1	5.4	5.8	6.1	6.4
150	2.9	3.6	4.1	4.6	5.0	5.4	5.8	6.2	6.5	6.8

Customs Form 6077 DOCK SHEET Entry Wh. 5072
Treasury Department SCIENTIFIC SPOT WEIGHING
March 1949

Port Wilmington Importer _____ Commodity Cigarette Leaf Tobacco

S.S. Malibu Arrived May 20, 1950 From Turkey Quantity 8000 Bales
U. S. Scales: Inspectors:

Customs Gross Pounds	Invoice Gross Pounds	Difference	Range	Mark & Bale No.	Customs Gross Pounds	Invoice Gross Pounds	Difference	Range
134.4	134.5	-.1		6724	138.5	138.3	.2	
139.5	139.4	.1		6166	136.9	136.7	.2	
135.7	135.6	.1		7754	133.5	132.8	.7	
138.5	138.3	.2		7099	134.6	134.5	.1	
134.0	133.9	.1	0.3	7044	133.4	133.4	.0	0.7
133.5	133.4	.1		7654	137.7	137.8	-.1	
135.4	135.6	-.2		7050	139.0	138.9	.1	
136.7	136.7	.0		7651	133.7	133.4	.3	
136.1	135.6	.5		7573	139.5	139.4	.1	
139.9	139.4	.5	0.7	7903	133.3	133.4	-.1	0.4
1363.7	1362.4				1360.1	1358.6		

	Summary		Range	
10 =	1337.5	1331.1	0.9	1.3
	1212.5	1208.4	1.7	0.7
	1319.5	1310.7	0.3	0.9
	1263.3	1256.8	1.1	1.1
	1231.9	1229.4	1.0	1.1
	1239.0	1235.8	0.7	1.5
	1262.0	1253.4	0.6	0.3
	1336.8	1326.1	0.8	0.7
	1363.7	1362.4	1.1	0.7
10 =	1360.1	1358.6	0.9	0.4
100 =	12926.3	12872.7		17.8 ÷ 20 = 0.89 = $\bar{R}_5$

CONVERSION FACTOR

12926.3 ÷ 12872.7 = 1.004163

RECAPITULATION

Invoice Gross		C. F.		Gross	Customs Tare	Net
1,031,477	x	1.004163	=	1,035,771	17,492	1,018,279

References

1. "ASTM Manual for the Presentation of Data," Table I, p. 50, 1947.

PROBLEMS OF THE EXPERIMENTER

W. J. Youden
National Bureau of Standards

Problems of the experimenter have had a controlling influence in the development of experimental design. In that sense the available experimental designs never quite meet the needs of today's experimenter. Most of us date the formal beginning of experimental design with the work of R. A. Fisher at Rothamsted beginning about 1920 (11). For many years agricultural field trials were severely handicapped by the heterogeneity of the soil. Efforts to improve the comparisons by increasing the number of plots were largely thwarted by the increased heterogeneity encountered as the area was made larger. The device of forming blocks, with the consequent reduction of the experimental error affecting the comparisons, made possible the detection of effects that otherwise would have been overlooked.

Inevitably the successes of designed experiments in agricultural field trials led to their application to other areas than agriculture. Soon it was discovered that agricultural field trials differed in two important respects from many other experiments. First, agricultural field trials, even with the device of blocks, usually involved fairly extensive replication. Second, the entire program was frozen at the start because all the plots were planted at one time and harvested together at a later time. Many experimenters had problems where the data inevitably came in a sequence and extensive replication was considered too costly or unnecessary. One development did carry over and that was the simultaneous investigation of several factors. A good deal was written about the wastefulness of investigations that studied the effect of one variable while holding all others constant. In point of fact, experimenters were studying multifactor problems long before the statistical concepts of main effects and interactions of factors came upon the scene. Statisticians contributed a new way of looking at the data obtained in multifactor experimentation and vastly elaborated the investigation of multifactor problems.

Again the requirement of running every possible combination that can be formed from several factors, each at several levels, brought new problems to the experimenter. Perusal of research journals will show figures with families of curves. Each curve represents the effect of varying one of the factors. The family of curves arises from varying some other factor or factors. Seldom are these curves nicely lined up so that the same range of the first factor is covered for the whole family of curves. Clearly there is here a departure from the notion of the complete factorial.

No one, therefore, should be surprised to find a new variety of experimental designs appearing upon the scene. The experimental observations may be viewed as defining a surface in two dimensions or in as many dimensions as there are factors. Statisticians therefore addressed themselves to devising designs that would facilitate the fitting of simple approximating functions that could be used to describe the results (2).

The large number of possible experimental combinations that accompany multifactor studies stimulated work in another direction. The well-known device of confounding high order interactions with blocks provided the key. Only one half, or a smaller fraction of these blocks, were actually used. Thus fractional replications (9) appeared and the experimenter was able to increase the number of factors investigated simultaneously. If the number of levels for each factor is held to two or three, these designs take on more of the character of a device for screening the most important factors from a number of factors.

Screening variables is an important problem for experimenters. In response to this need, experimental design has progressed from the more intensive study of a few factors to include studies that are extensive and selective. The argument is made that once the important variables have been isolated they may be studied more intensively at the option of the experimenter.

We are now in the midst of a good deal of activity in the development of screening programs. More than twenty years ago Yates (18), in a paper on complex experiments, referred to this problem. He pointed out that seven objects could be weighed in various combinations so that in just eight weighings the differences between the means of four weighings could be used to estimate each weight. These estimates would have one half the standard error of individual weighings. Yates further remarked: "Cases in which interactions are certainly negligible are, in fact, rather rare. Even in the example given above, although weights are undoubtedly additive, systematic errors of the apparatus are likely to complicate the issue. But such experimental systems may be useful in certain preliminary surveys, where there is good reason to believe that interactions are small and where it is desired to include as many factors as possible (some of them perhaps unlikely to produce any effect whatsoever)."

Preliminary surveys today loom more important than twenty years ago. Much has been heard of random balance as a technique for exploring the effects of a score or more variables using as few as thirty or thereabout observations(3). The method of prescribing the level of each of the twenty variables for a given run is disarmingly simple. The level is left to chance for each variable subject to the restriction that, over the whole number of runs, each level is used the same, or

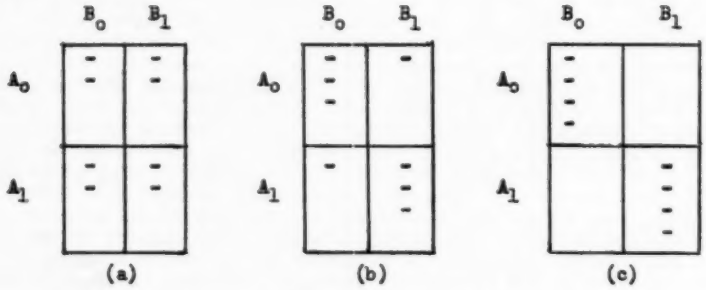
nearly the same number of times. This technique of programming the runs certainly is easier than working out fractional factorials. Indeed a slip in the factorial specifications means an awkward adjustment in the analysis and would appear not to matter at all in a random balance program. A good many problems await solution in connection with the application of this technique of experimentation. There are indications that a closer control over the assignment of the levels of the variables is both possible and much to be desired.

The foregoing remarks constitute a terse sketch of the undoubted progress that has been made towards meeting the problems of the experimenter. Experimental designs require whole text books for their cataloging and analysis (4, 5, 6, 8, 10, 12, 13, 14, 16, 17). The experimenter is confronted with the problem, not of selecting which variables to study, but which designs to use. Of course, his statistical consultant is supposed to solve this problem for him too. Obviously a good selection depends upon how successfully the statistical consultant has grasped the nature of the experimental problem that is at hand. The choice of design, too, will probably bear some correlation with the background of the consultant. Some light on the background may be had by finding out if the consultant has contributed to the literature of fractional factorials or to surface fitting or to random balance. There was a time, when experimental designs were few in number, that it was often necessary to alter the experimenter's problem so that it would fit some available design. One measure of our progress in experimental design is the reduction in the amount of changing of the problem that is now necessary. We ought to be very careful here or perhaps we may solve a problem but not the experimenter's problem.

There are many varieties of experimental design but fortunately there are only a few basic items that constitute a sort of check list of the characteristic signs of good experimentation. Experimenters themselves established most of the items on the list. There seems to me some danger that experimenters may believe that all experimental designs have all the good characteristics. The experimenter cannot evade his final responsibility for the conduct of the experiment. Most technicians can tune up the apparatus and run through a procedure of measurement more skillfully than the man charged with the responsibility for the investigation. All the experimental designs now and hereafter available will not relieve the investigater of his responsibility for the project. For this reason, the experimenter must make sure, before he irretrievably embarks on a program, that he understands what he can expect to get and what he cannot expect to get from a particular variety of experimental design.

One of the desirable characteristics in experimentation is high efficiency. We may illustrate this concept by considering how eight measurements may be allocated between two items, A and B, that are to be compared. Four measurements on each item is the best possible division of the effort. One measurement on A and seven on B will give a less precise comparison than four measurements split equally.

Consider eight measurements available to study two variables each at two levels. There are three ways of assigning the measurements to the four combinations: A and B both at low level; A and B both at high level; A high and B low; and A low and B high. A sure instinct for balance guides us to assign two measurements to each combination as in (a).



In the extreme case, shown in (c), the individual effects of the variables can not be determined.

The random assignment of two levels for each of two variables gave the following examples of unequal balance when 32 runs were made. Thirty repetitions were tried and an exact balance, with eight of each combination, was found in only three cases.

	4	12		5	11		6	10		7	9		8	8
	12	4		11	5		10	6		9	7		8	8
frequency		1			1			7			18			3

Random balance programming of the experiment confounds main effects to varying degrees and this loss of efficiency is part of the price paid for including a very large number of variables.

Among the trying tasks that sometimes come to a statistician is the evaluation of results in which the experimental program left a choice of ways for operating upon the data. One of the great benefits in the early history of experimental design was the completely unambigious specification of the arithmetic to be performed upon the data. This specification was drawn up before the data were in hand. Modern advances in statistics allow some relaxation of this prior formulation of the analysis. It can be said that a price, usually in wider confidence limits, is exacted for the priviledge of being guided by the data in the analysis (7, 15).

Probably the greatest advance in the improvement of experimentation came with the use of randomization as a means of insuring the validity of the estimate of the experimental error. The careful experimenter may use either a t-test or his hard bought experience as a yardstick for appraising his data. A classic example dates back before Student's t-test was available. Lord Raleigh was led to the discovery of the rare gases in the atmosphere because the difference he found between the densities of nitrogen prepared from the atmosphere and from chemical compounds exceeded the difference he considered likely to arise from the experimental error in the determinations.

The estimate of error for fractional factorial designs uses the contrasts assigned to high order interactions. In so far as these interactions are not zero, the estimate of the error is inflated and the chance of detecting effects correspondingly reduced. No method has been described for estimating the experimental error in random balance experiments. Perhaps this is not necessary in a screening process. The screening procedures studied by Bechhofer and his associates (1) do give the experimenter a precise estimate of the chance that an effect of any specified magnitude will not be overlooked, if in fact present.

No one disputes the complexity of the situations that confront the experimenter. The experimenter has his work cut out for him any way you look at it. Past experience often makes it possible for an experimenter to rate in order of importance the variables that abound in his problem. Customarily a fairly limited number from the top of the list are taken for the study. What about the other variables? All experimenters have learned the hard way that if these other variables are allowed to vary excessively, their effects pile up in the experimental error and obscure the effects of the variables under study. Confronted with these nuisance variables, the experimenter makes considerable efforts to hold them constant for the duration of the study. The device of local control is often possible also. All factors, not specifically a part of the study, (including any

unidentified factors), are hopefully considered to be more constant for a short interval than over a long interval. The total program is therefore divided into a number of periods, or blocks, and the runs carefully alloted to the various blocks. In this way, much of the disturbing effects of these variables may be removed as block effects with a corresponding improvement in the power to detect effects among the variables actually studied (13).

On the chance that one or more really important variable may be hidden among this list formerly held as nearly constant as possible, some form of random balance experimentation may be considered. In this event these variables are purposely varied. Many of these variables whose effects fall short of detection will be partially confounded with other variables and will reduce the discriminatory power of the experiment.

The experimenter must remember that there is no one technique of experimentation that will be best in all circumstances. Unquestionably, a technique that greatly enlarges the number of variables studied must do so at the expense of other aspects of the experiment, and in particular, some aspects to which the experimenter has come to expect to be present. The experimenter, to avoid disappointment, must be careful to ascertain just what a particular type of design can be expected to give him. There are enough disappointments connected with experimentation without inviting others.

Literature Cited

(1) Bechhofer, R. E., A single-sample multiple decision procedure for ranking means of normal populations with known variances. Ann. Math. Stat. 25, 16, (1954).

(2) Box, G. E. P., and Hunter, J. S., Multifactor experimental designs. Ann. Math. Stat. 195, 28, (1957).

(3) Budne, T. A., SQC can be made more effective. Industrial Quality Control 15, 10, (1958).

(4) Cochran, W. G. and Cox, G. M., Experimental designs, 2nd Ed. John Wiley and Sons, New York, (1957).

(5) Cox, D. R., Planning of experiments. John Wiley and Sons, New York, (1958).

(6) Davies, O. L., Design and analysis of industrial experiments. Oliver and Boyd, London, (1954).

(7) Duncan, D. B., Multiple range and multiple F tests. Biometrics 11, 1, (1955).

(8) Federer, W. T. Experimental design, theory and application. The Macmillian Co., New York, (1955).

(9) Finney, O. J., The fractional replication of factorial arrangements. Ann. Eugen. 12, 291-301, (1945).

(10) Finney. O. J., Experimental designs and its statistical basis. Univ. Chicago Press., Chicago, (1955).

(11) Fisher, R. A., Statistical Methods for Research Workers, 1st Ed., Oliver and Boyd, Edinburgh, (1925).

(12) Fisher, R. A., The design of experiments, 5th Ed., Oliver and Boyd, Edinburgh, (1949).

(13) Fractional factorial experiment designs for factors at two levels, Applied Math. Series 48, Government Printing Office, Washington. (1956).

(14) Kempthorne, O., The design and analysis of experiments. John Wiley and Sons, New York, (1952).

(15) Scheffe', H., A method for judging all contrasts in the analysis of variance. Biometrika 40, 87, (1953).

(16) Quenouille, M. H., The design and analysis of experiment. Hafner Publishing Co., New York, (1953).

(17) Villars, D. S., Statistical design and analysis of experiments for development research. Wm. C. Brown Co., Dubuque, Iowa, (1951).

(18) Yates, F., Complex experiments. Supplement Jour. Royal Statistical Society 2, 211, (1935).

QUALITY CONTROL APPLICATIONS IN STAMPINGS, WELDING AND THE JOB SHOP

J. Mehalek
The Budd Company

A discussion of the methods used in the control of stampings and welding in automobile body construction and job shop operations. This encompasses the preparation for manufacture and production of a variety of products usually made by a stamping plant. These can best be divided into two groups:

1. Basic Controls
2. Production Controls

Basic controls cover engineering, tooling, preproduction, initial production, with the many methods devised for assuring conformance with the requirements of the end product. Production controls are on production and include such items as tool control, first piece inspection, welding control, non-destructive testing and quality audit systems.

Engineering Controls consists of checking of drafts and models prior to working them into the system for the purpose of making tools to produce the product.

Tooling Controls are comprised of close checks on all tools to assure satisfactory performance. This includes inspection fixtures, gages, templates, etc.

Preproduction Controls require the careful detailing and study of all procedures and processes to assure satisfactory results. Included in this are die tryouts, proper functioning of assembly fixtures and development of weld controls (qualification of welding equipment and schedules).

Initial Production Control includes first piece inspection on stampings and assemblies, by a master checking unit setting up and laying out the details. Also by a complete try-out with the use of all tools and inspection facilities.

Production Controls consist of detailed inspections, surveillance checks on tools and product, welding control through periodic maintenance of welding equipment developed from records of previous performances. The correlation of customer reports.

Control through Non-destructive Tests. In the case of The Budd Company, we are continually working to develop improved methods of manufacturing and quality control, particularly through non-destructive tests. Some of the newer processes recently developed or still under development are:

The Monautronic Welding Control, a means of assuring satisfactory welds, through a resistance welding control system which guarantees fusion at the interface by a technique called "Voltage Constraint".

Gammascan, a means for measuring the thickness of metals and other materials by use of radioactive sources and scintillating crystals, very accurate and very fast.

Portable Radiography using Cobalt 60 and Iridium 192, ideal for taking X-ray pictures in normally inaccessible places.

Permaweld a means of checking resistance welds. This is a non-destructive test for weld assurance making a permanent record of the weld which is reproduceable either by ozalid or blue print processes.

Control through Quality Audit Systems have a wide range of application and are being used quite extensively by some of the larger companies. Using detailed check-off sheets, a quality audit group will select a random sample of the product and proceed to make a very detailed check of the quality, recording their findings on the check sheets. These records are kept from audit to audit and used in plotting control levels and operating characteristic curves. Those using this system, particularly where they are confronted with a multiplicity of problems on a complex assembly, have found the quality audit system to be the best management tool for assigning priorities and directing action.

Control Records on all the above controls are carefully recorded, these records being used for reference on subsequent runs and for information as required by other departments such as sales, engineering, manufacturing, purchasing, etc. The records are also the basis for keeping management informed as to the quality standards of the product; statistics being compiled on acceptances, charges, costs, scrap, etc.

RANDOMIZED INCOMPLETE BLOCKS

Charles R. Hicks
Purdue University

I believe I have been asked to talk on Incomplete Block Designs because there seems to be few, if any, papers on this subject either in Industrial Quality Control or in A. S. Q. C. Convention Proceedings. It seems safe to assume, therefore, that the audience is not familiar with these designs and an attempt should be made to carefully define our terms and illustrate the design with a numerical example.

Many quality control people are quite proficient at techniques such as testing hypothesis, analysis of variance, multiple regression, etc., but often unaware of the basic design of the experiment whose results are analyzed by these and other techniques. The design of any experiment should include provision for the analysis to be made on the resultant data but should also include the details of just how the experiment is to be run. It should include the method of randomization to be used in collecting the data. In many plant situations it is impractical and often impossible to completely randomize the order in which samples are to be run. Such experimental designs as randomized blocks, Latin squares, incomplete blocks, Youden squares, etc., represent restrictions on the randomization procedure. Such restrictions are often beneficial in isolating important sources of variability in the experiment.

In order to define incomplete blocks, one first must define a complete block. Let us consider an experiment in which interest centered on the effect of four treatments applied to the coils of TV tube filaments on the current flow through these coils. As the treatments required time to apply to the coils, it was decided to run the four treatments, say, A, B, C, and D, on 4 different days making 16 experiments in all. Now, if the treatments are selected completely at random for the four days with only the stipulation that each treatment is to be used four times, the order of experimentation might be as follows:

	Treatments			
Day 1	D	C	B	C
Day 2	D	A	C	B
Day 3	B	D	B	C
Day 4	A	D	A	A

Table 1

Such a design shows that certain days included no experiment with treatment A, for example. If any difference in current flow occurred on days 1 and 3 from days 2 and 4, it might show up as an apparent effect of treatment A rather than a day-to-day effect. Also, as such designs use the reproducibility of current flow under the same treatment as a yardstick for judging the treatment effect such a yardstick or error variance may be highly inflated by the day-to-day variation if present.

For these reasons, a randomized block design might be used with days constituting the blocks. A block is defined as a set of observations in which the error variance is expected to be less than in the whole series of observations. For this problem, we might have:

Day	Treatments				
1	C	B	D	A
2	B	A	C	D
3	C	A	D	B
4	C	B	D	A

Table 2

Here the order of treatments is randomized within each block with the restriction that each treatment appears once and only once in each block. Such a design may be analyzed as a 2-way classification analysis of variance using the model:

$$X_{ij} = m + t_i + b_j + e_{ij}$$

where m is the common effect, t_i the treatment effect, b_j the block effect (here days) and e_{ij} the random error. Such designs assume no treatment by block interaction as it would be confounded with the error. However, if blocks are considered as a random effect and treatments fixed, it is not necessary to assume no interaction as a valid test on the treatments is still available if treatments are the main concern of the experimenter. By removing the block effect (days) from the data, a smaller error term should result and give a good test on the effects of treatments on the current flow.

In some experiments it is impossible to try all treatments in every block. In such a case, we have an incomplete block. For example, in the problem above, if only 3 treatments could be run each day and we were interested in studying four treatments, the blocks would be incomplete having only 3 treatments per block rather than four. In order to get valid comparisons for all four treatments, we usually choose enough blocks (days) so that each pair of treatments appear together the same number of times in the blocks. This design is called a balanced incomplete block design. Results on current flow measurements (coded by subtracting 513 ma from all readings) are displayed below:

Days (blocks)	Treatment A	B	C	D	B_j
1	2	-	20	7	29
2	-	32	14	3	49
3	4	13	31	-	48
4	0	23	-	11	34
T_i	6	68	65	21	160=T

Table 3

In table 3, on the first day, treatments A, C, and D were run and results recorded. On day two, B, C, and D were run. Here we had 3 treatments per block, 4 treatments in all, 4 blocks and 3 replications of each treatment (run on 3 different days). Also each pair of treatments appear together twice in the blocks. Here again the order in which the 3 treatments were run each day would be randomized.

In order to analyze these data in this incomplete block design, we first need some notation: Let:

b = number of blocks, here b=4
t = number of treatments, here t=4
k = number of treatments per block, here k=3
r = number of replications of each treatment, here r=3
λ = number of times each pair of treatments occur together in the blocks, here λ=2

Note, if N=total number of observations,

$$N = bk = tr, \text{ here } N=12$$

The Analysis begins in the usual manner for an analysis of variance. Calculate the Total Sum of Squares:

$$SS_{Total} = \sum_i \sum_j X_{ij}^2 - \frac{T^2}{N} = 3478 - \frac{(160)^2}{12} = 1344.67$$

Calculate the block effect, ignoring treatments

$$SS_B = \sum_j \frac{B_j^2}{k} - \frac{T^2}{N} = 2234 - \frac{(160)^2}{12} = 100.67$$

To calculate the treatment effect adjusted for blocks, we first compute the quantities we will call Q_i, where

$$Q_i = k\,T_i - \sum_j n_{ij}\,B_j$$

here $n_{ij} = 1$ if treatment i appears in block j, and $n_{ij} = 0$ if treatment i does not appear in block j. This $\sum_j n_{ij}\,B_j$ is merely the sum of all blocks which contain treatment i.

For our data: $Q_1 = 3\,(6) - 111 = -93$

$Q_2 = 3\,(68) - 131 = 73$

$Q_3 = 3\,(65) - 126 = 69$

$Q_4 = 3\,(21) - 112 = -49$

Note that $\sum_i Q_i = 0$. From these Q values, the Sum of Squares due to treatments adjusted for blocks is given by:

$$SS_{T(adj.)} = \frac{\sum_i Q_i^2}{k\lambda t}$$

Here: $SS_{T(adj.)} = \frac{(-93)^2 + (73)^2 + (69)^2 + (-49)^2}{(3)(2)(4)} = 880.83$

Summarizing in an ANOVA table:

Source of Variation	d.f.	S.S.	M.S.	E.M.S.
Treatments (adj.)	3	880.83	293.61	$\sigma^2 + \frac{8}{3}\sigma_T^2$
Blocks	3	100.67	33.56	discard.
Error (by subtraction)	5	363.17	72.63	σ^2
Total	12	1344.67		

Table 4

The F test yields $F_{3,5} = \frac{293.61}{72.63} = 4.04$ which is not significant at the 5% level ($F_{3,5,.05} = 5.4$). We conclude that these four treatments show no significant effect on current flow at the 5% level.

The EMS column in general carries the coefficient on σ_T^2 of $r E_f$ where E_f is defined as the efficiency of the design compared to a complete block design. This efficiency can be shown to be equal to $\frac{\lambda t}{kr}$. Here $E_f = \frac{2(4)}{3(3)} = \frac{8}{9}$ or this design is approximately 89% as efficient as a complete block design. If complete the coefficient on σ_T^2 would be the number of replication per treatment or r. Here it is $r E_f = 3(\frac{8}{9}) = \frac{8}{3}$

If one wishes to recover block effects, it can be done in the same manner, provided, the design is symmetric, i.e. t=b.

As before

$$SS_{Total} = 1344.67$$

$SS_T = 3108.67 - 2133.33 = 975.34$

Blocks. $Q'_j = r B_j - \sum_i n_{ij} T_i$

$Q'_1 = 3(29) - 92 = -5$

$Q'_2 = 3(49) - 154 = -7$

$Q'_3 = 3(48) - 139 = 5$

$Q'_4 = 3(34) - 95 = \frac{7}{0}$

$$SS_{B(\text{adj. for trts.})} = \frac{\sum_j Q'^2_j}{r\lambda b} = \frac{\sum_j Q'^2_j}{24} = 6.17$$

Complete ANOVA:

	d.f.	S.S.	M.S.	E.M.S.
Bet. Trts. (adj.)	3	880.83	293.61	$\sigma^2 + \frac{8}{3}\sigma^2_\gamma$
Bet. Blocks	(3)	(100.67)	discard	
Bet. Blocks (adj.)	3	6.17	2.06	$\sigma^2 + \frac{8}{3}\sigma^2_\beta$
Bet. Trts.	(3)	(975.34)	discard	
Error	5	363.17	72.63	σ^2
Total	11	1344.67		

Table 5

Note: adj S.S. do not add to Total S.S. as the design is basically non-orthogonal.

Here neither treatments nor blocks are significant at the 5% level.

An alternative method may be used to arrive at the same results as given in table 5. This method is more general than the one given above as it can be used whether the design is symmetric or not and can even be used when the blocks are unbalanced. In general, however, this alternative method is much longer than the one already described. We will illustrate it with the same data.

The first step is to complete the incomplete block design by filling in missing values. The method is to fill in values (α, β, γ, δ) which would make the S.S. for the error a minimum:

		Treatment				
		A	B	C	D	
Block	1	2	α	20	7	$29+\alpha$
	2	β	32	14	3	$49+\beta$
	3	4	13	31	γ	$48+\gamma$
	4	0	23	δ	11	$34+\delta$
		$6+\beta$	$68+\alpha$	$65+\delta$	$21+\gamma$	$160+\alpha+\beta+\gamma+\delta$

Table 6

$$\text{S.S. error} = \alpha^2 + \beta^2 + \gamma^2 + \delta^2 + 2^2 + \ldots + 11^2 - \frac{1}{4}\left[(6+\beta)^2 + (68+\alpha)^2 + (65+\delta)^2 + (21+\gamma)^2 \right] - \frac{1}{4}\left[(29+\alpha)^2 + (49+\beta)^2 + (48+\gamma)^2 + (34+\delta)^2 \right] + \frac{1}{16}(160 + \alpha + \beta + \gamma + \delta)^2$$

If this expression is differentiated with respect to α, β, γ and δ to minimize the error S.S. and set equal to zero, the following equations result:

$$9\alpha + \beta + \gamma + \delta = 228$$

$$\alpha + 9\beta + \gamma + \delta = 60$$

$$\alpha + \beta + 9\gamma + \delta = 116$$

$$\alpha + \beta + \gamma + 9\delta = 236$$

If these four equations are now solved simultaneously for α, β, γ and δ by the Abbreviated Doolittle or some other method, the resulting missing values are:

$$\alpha = 21.8$$

$$\beta = 7.8$$

$$\gamma = 0.8$$

$$\delta = 22.8$$

These values are now substituted back into the design and a straight two-way ANOVA is run for the sole purpose of recovering the error S.S. Recall that this is a minimum for the values above. For these data, S.S. error = 363.17. Note that this is the same as the error S.S. in tables 4 and 5.

Next, a one-way ANOVA is performed on treatments ignoring blocks and leaving out the missing values. The result yields $SS_{Trts.} = 975.34$ and $SS_{Total} = 1344.67$. The difference 369.33 must therefore contain the random error and the block effect adjusted for treatments. But we know the random error SS = 363.17, hence the difference 369.33 - 363.17 = 6.16 must be SS for blocks adjusted for treatments. This is the same value as given in table 5 within the error of calculations.

If a similar procedure is carried through for blocks ignoring treatments, the treatment effect SS adjusted for blocks is recovered and equals 880.83 which checks with table 5.

This second method seems to involve much more work and could be quite messy if there are very many missing values. Its generality may, nevertheless, make it useful in many situations.

In conclusion, we might say that the incomplete block design is very useful when complete blocks cannot be formed. The design is, of course, less efficient than complete blocks and in practice it may be quite costly to run enough blocks to provide a balanced incomplete block design. Many textbooks give tables for typical balanced incomplete block designs (4, 5).

Appendix

The mathematical model for an incomplete block design is given by:

$$X_{ij} = n_{ij}(\mu + \tau_i + \beta_j + \varepsilon_{ij})$$

where

$n_{ij} = 0$ if the ith treatment does not appear in the jth block.

$= 1$ if the ith treatment does appear in the jth block.

Here $i = 1, 2, \ldots t$ treatments

$j = 1, 2, \ldots b$ blocks

and $\sum_{j=1}^{b} n_{ij} = r$ reps. per trt.

$\sum_{i=1}^{t} n_{ij} = k$ trts. per block

The sample model would be:

$$X_{ij} = n_{ij}(m + t_i + b_j + e_{ij})$$

The sum of the squares of the errors $\sum_{i,j} (n_{ij}\, e_{ij})^2$ would be:

$$\sum_i \sum_j (X_{ij} - n_{ij}\, m - n_{ij}\, t_i - n_{ij}\, b_j)^2$$

Differentiating with respect to m:

$$-2 \sum_i \sum_j (X_{ij} - n_{ij}\, m - n_{ij}\, t_i - n_{ij}\, b_j)\, n_{ij} = 0$$

$$\sum_i \sum_j n_{ij}\, X_{ij} - m \sum_i \sum_j n_{ij}^2 - \sum_i \sum_j n_{ij}^2\, t_i$$

$$- \sum_i \sum_j n_{ij}^2\, b_j = 0$$

But $\sum_i \sum_j n_{ij}^2 = \sum_i \sum_j n_{ij}$ as n_{ij}'s are either 0's or 1's,

and $\sum_i \sum_j n_{ij} = N$. Also $\sum_i \sum_j n_{ij}\, X_{ij} = T$ (Total of

observations) $T - m\, N - \sum_i t_i \left(\sum_j n_{ij}^2\right) - \sum_j b_j \left(\sum_i n_{ij}^2\right) = 0$

But $\sum_j n_{ij}^2 = \sum_j n_{ij} = r$ and $\sum_i t_i = 0$

and $\sum_i n_{ij}^2 = \sum_i n_{ij} = k$ $\sum_j b_j = 0$

The equation then reduces to

$$m\, N = T \quad \ldots\ldots (1)$$

Differentiating with respect to t_j :

$$\frac{\partial}{\partial t_i} \sum_j (X_{ij} - n_{ij}\, m - n_{ij}\, t_i - n_{ij}\, b_j)^2$$

$$- 2 \sum_j (X_{ij} - n_{ij}\, m - n_{ij}\, t_i - n_{ij}\, b_j)\, n_{ij} = 0$$

$$\sum_j n_{ij} X_{ij} - m \sum_j n_{ij}^2 - t_i \sum_j n_{ij}^2 - \sum_j n_{ij}^2 b_j = 0$$

$$T_i - m\,r - t_i\,r - \sum_j n_{ij}\, b_j = 0$$

$$\text{Or, } m\,r + r\,t_i + \sum_j n_{ij}\, b_j = T_i \quad \ldots\ldots(2)$$

Differentiating with respect to b_j :

$$\frac{\partial}{\partial b_j} \sum_i (X_{ij} - n_{ij}\, m - n_{ij}\, t_i - n_{ij}\, b_j)^2$$

$$-2 \sum_i (X_{ij} - n_{ij}\, m - n_{ij}\, t_i - n_{ij}\, b_j)\; n_{ij} = 0$$

$$\underbrace{\sum_i n_{ij} X_{ij}}_{B_j} - m\,k - \sum_i n_{ij}\, t_i - b_j\, k = 0$$

$$\text{Or, } m\,k + \sum_i n_{ij}\, t_i + k\, b_j = B_j \quad \ldots\ldots(3)$$

The equations numbered 1, 2 and 3 above actually represent $1 + t + b$ equations to be solved for m, b_j's and t_i's.

Equation 2 cannot be solved directly for the treatment effect t_i as it is corrupted by the block effects b_j. In a complete block all $n_{ij} = 1$ and then $\sum_j n_{ij}\, b_j = \sum_j b_j = 0$ and t_i <u>can</u> be solved for directly and independent of block effects. Such is the nature of orthogonal designs.

To recover treatment effect t_i here, divide equation 3 by k, multiply it by n_{ij}, sum over all blocks and subtract from equation 2:

$$\div 3 \text{ by k:} \quad m + \frac{1}{k} \sum_i n_{ij}\, t_i + b_j = \frac{B_j}{k}$$

times n_{ij} and $\sum_j$: $\quad m \sum_j n_{ij} + \frac{1}{k} \sum_\ell \sum_j n_{\ell j}\, n_{ij}\, t_i$

$+ \sum_j n_{ij}\, b_j = \sum_j \frac{n_{ij} B_j}{k}$

This term is $n_{\ell j}$ as it may be a different n than for the ith treatment.

Subtracting from (2) $\quad m r + r t_i + \sum_j n_{ij}\, b_j = T_i$

$$m r + \frac{1}{k} \sum_\ell \sum_j n_{\ell j}\, n_{ij}\, t_i + \sum_j n_{ij}\, b_j = \sum_j \frac{n_{ij} B_j}{k}$$

$$r t_i - \frac{1}{k} \sum_\ell \sum_j n_{\ell j}\, n_{ij}\, t_i = T_i - \sum_j \frac{n_{ij} B_j}{k} \quad \ldots.(4)$$

If the term on the right of 4 is multiplied by k it will equal the Q_i defined in the paper, i.e.

$$r t_i - \frac{1}{k} \sum_\ell \sum_j n_{\ell j}\, n_{ij}\, t_i = \frac{Q_i}{k} \quad \ldots.(5)$$

In a balanced incomplete block, λ is the number of times each treatment pair appear together. Now

$$\sum_j n_{\ell j}\, n_{ij} = \begin{cases} \lambda \text{ if } i \neq \ell \\ r \text{ if } i = \ell \end{cases}$$

$$r t_i - \frac{1}{k} \left(\sum_j \left[n_{\ell j}\, n_{ij}\, t_i + \sum_{i \neq \ell} n_{\ell j}\, n_{ij}\, t_\ell \right] \right) = \frac{Q_i}{k}$$

where $i = \ell$

$$r t_i - \frac{1}{k} \left(r t_i + \lambda \sum_{\ell \neq i} t_\ell \right) = \frac{Q_i}{k}$$

$$r t_i - \frac{1}{k} \left(r t_i + \lambda \left[\sum_{i=1}^{t} t_i - t_i \right] \right) = \frac{Q_i}{k}$$

But $\quad \sum_i t_i = 0$

$$t_i \left(r - \frac{r}{k} + \frac{\lambda}{k}\right) = \frac{Q_i}{k}$$

$$\left[\frac{r(k-1) + \lambda}{k}\right] t_i = \frac{Q_i}{k} \quad \ldots\ldots(6)$$

Hence the treatment effect is equal to Q_i/k divided by a constant. Let

$$r E_f = \frac{r(k-1) + \lambda}{k}$$

Then $$t_i = \frac{Q_i}{k\, r\, E_f}$$

As this t_i is a regression coefficient in the model, the SS due to t_i or due to treatments adjusted for blocks is:

$$SS_{T(adj.)} = \sum_i \frac{Q_i^2}{k^2\, r\, E_f}$$

But $$r E_f = \frac{r(k-1) + \lambda}{k}$$

$$SS_{T(adj.)} = \frac{\sum_i Q_i^2}{\left[k^2 \frac{(r(k-1) + \lambda)}{k}\right]} = \frac{\sum_i Q_i^2}{k\left[r(k-1) + \lambda\right]}$$

As there are r choices of the 1st trt. to go into a block and after that k-1 out of t-1 choices to pair with this treatment,

$$\lambda = r\,\frac{(k-1)}{t-1}, \text{ hence } r(k-1) = \lambda(t-1)$$

$$SS_{T(adj.)} = \frac{\sum Q_i^2}{k\left[\lambda(t-1) + \lambda\right]} = \frac{\sum_{i=1} Q_i^2}{k\lambda t}$$

as used in the example.

References

1. Anderson, R. L. and Bancroft, T. A. "Statistical Theory in Research", McGraw-Hill, 1952

2. Bennett, C. A. and Franklin, N. L. "Statistical Analysis in Chemistry and the Chemical Industry", Wiley 1954

3. Chew, V. "Experimental Designs in Industry", Wiley, 1958

4. Cochran, W. G. and Cox, G. M. "Experimental Designs", Wiley, 1950

5. Davies, O. L. "Design and Analysis of Industrial Experiments", Hafner, 1954

THE ROLE OF PRODUCT ENGINEERING IN AUTOMOTIVE QUALITY

John R. Parker
A. O. Smith Corporation

In reviewing the role of Product Engineering in automotive quality, and in attempting to highlight things that may be useful, generally, for quality improvement, we are directing this discussion toward the professional and management people in the Quality Control organization. We will concentrate on the relationship between these two functions whose responsibilities are so closely allied with the thought that any improvement which can be achieved in cooperative effort between Engineering and Quality Control will be helpful to those addressed.

Since my background includes Product Engineering experience and more recently Quality Control as well, I possibly see this interrelationship from a somewhat different viewpoint. I hope that this viewpoint can provide a little novelty which may help you in your continuing effort to sell the importance of quality thinking to the members of your management group and perhaps more effectively to enlist the help of your own Product Engineering group.

Every part of the manufacturing organization is important to the quality of the end product. Undoubtedly, one of the most important activities of the Quality Control Management is the continuing promotion of this fact and the highlighting, for each of the other functional managers, the areas in which their actions effect quality performance and the ways in which they can contribute to better quality.

In the case of Product Engineering, there are some very obvious responsibilities to the quality result and there are also some that are less obvious. In most businesses, it is the Engineering Department which is charged with the responsibility for evaluation of the product characteristics that the customer wants and for the definition of these characteristics and translation of them into a set of performance specifications. This is true to a large extent in the automotive industry although stylists, product planners, and others share in this responsibility. So, if we think of quality in broad terms, that is if we define a product of high quality as one which possesses a maximum of those characteristics which contribute to its desirability from the customer's viewpoint, then this definition of the desired end result, this establishment of the goal, is Engineering's key quality function. More specifically, the product engineer must translate these desired characteristics into detail physical specifications of the things being manufactured, and this again is a quality function, since the customer's judgment of product quality includes both the adequacy of the specifications and the degree of attainment.

These are the two obvious areas in which the product engineer enters the quality picture; obvious because they are a part of his job, by definition. There is a corollary responsibility which really isn't part of the quality picture at all, but which has such an important bearing on it that it must be mentioned here. This is the engineer's responsibility for product cost; as we shall see, it demands a fair

share, with quality, of the engineer's attention. Performance is hardly meaningful except in relation to its price tag. Modern technology being what it is, it would literally be possible to achieve any level of performance desired if sufficient time, labor and material were committed. So Engineering must continually be evaluating whether each performance or appearance benefit is worth its cost. I don't mean to imply that quality improvements invariably add cost. There are many examples of design improvements that have actually reduced costs, and very often better quality performance in manufacturing is accompanied by reduced costs; but the usual Engineering decision is between product improvement and cost.

In any event, ideal quality is that which is sufficiently good for the purpose intended and is achieved at a minimum of cost. Product Engineering's contribution to quality must be evaluated, then, within the framework of the cost designed into the product. Quality managers might be helpful, I believe, by politely reminding their Engineering Department friends of the fact that intangible costs count, too, and that extra inspection and repair operations are as often the result of poor engineering as of poor manufacturing.

So far then, we have said that the control of anything, whether it be product quality, cost, or schedule, implies the establishment of a standard or goal and the application of tests to determine if the goal has been achieved; and that against this definition, Product Engineering's obvious role in the quality picture is that of the establishment of the goal and the translation of it into detail product requirements. We have also said that, from a practical standpoint, the performance of these engineering functions can only be judged in reference to the cost involved in achieving the standards set.

There are a number of less obvious Product Engineering involvements in quality, and it is some of these areas that the interrelationships between Quality Control and Product Engineering as departments in a manufacturing concern can be most productive.

We have said that the Engineering Department must translate overall performance requirements to detail product specifications. This constitutes the biggest and most time consuming share of the work of the Engineering group as it involves, theoretically at least, the definition of every dimensional, visual, and material composition characteristic of each part and each sub-assembly such that the final assembly will perform as specified and at minimum cost. The Quality Control Department can work hand in glove with this effort since it is the feed back system which tells whether or not detail specifications are met and which provides the necessary data and analysis to identify reasons for failing to meet these characteristics.

In addition, as a result of experience on similar prior models, the Quality Department is usually a very authorative source for evaluating the detailed descriptions of the various elements of the product with respect to their adequacy as fabricating instructions and also with respect to the actual requirements of the buyer and the user. In short, the Quality Department can be most helpful to the Engineering Department on this very important phase of their work and I think that

both Engineering and Quality Control Departments should be aware of this fact and that neither one should be reticent about taking the initiative to use and improve on this capability.

Obtaining the proper relationship here is largely a matter of communications; the statistical methods of the modern Quality Control organization are, in my opinion, about as important in their contribution to communications between Quality Control and Engineering as in any other respect. The engineer is trained to think quantitatively and he reacts best to quantitative salesmanship. That is to say, that his thinking can be influenced most effectively when it is possible to provide quantitative data. The data which the Quality Control organization has now prepared to provide makes it more effective in reporting conformance to established specifications and makes it many, many more times effective in contributing to the initial definition of the specifications themselves.

In the old days when an inspector expressed an opinion concerning the necessary or practical degree of restriction of a specification, the engineer had no choice but to view it as just what it was--an informed opinion, but nevertheless, an opinion. But when today's Quality Control Department can show quantitative statistical data, fact has been substituted for opinion. Usually, the result is that another engineer is on the Quality Control band wagon and, when it comes time to work on next year's model, he will go back to the proven source for authoritative, reliable information on which to base his design.

In addition to establishing detail characteristics, the Engineering Department, through experience with the preliminary development testing, is often the best source for the establishment of performance tests. Here, too, Quality Control can be effective, using its background of experience with similar tests, in guiding Engineering on the practical necessities for successful shop tests.

Often the Engineering Department suggests or helps to develop new processing methods which are required for a new product design. This familiarity with the important aspects of the processing methods makes Engineering a potential authority for consultation in the establishment of shop controls for the process after it goes into production. This, then, is an area in which Quality Control can use the Product Engineering Department as a help in the performance of its own task.

Another area in which Product Engineering can supplement the Quality Control program is one which has become increasingly apparent to me over the past few years. Engineering can serve as an ideal training ground for at least some portion of the Quality Control personnel. In the first place, the rapid development of the technical requirements for Quality Control work which has occurred in recent years, makes an engineering education very desirable for Quality Control supervisory personnel. Of course, the largest supply of such persons in most companies is in the Engineering Department, and generally speaking, the best organized program for recruiting and giving preliminary training to these people is the same area. Furthermore, experience with the requirements of the product from an engineering standpoint and experience with the processing methods employed and

problems encountered are both obtained by the aggressive Product Engineering employee and usefully applied in Quality Control. The results of our addition of a few Engineering trained people to our Quality Control organization has been very encouraging.

We said earlier that one of Quality Control's main jobs is continually to educate the balance of the organization in the appreciation of, and the means for correction of, quality problems. I feel that the value of the Product Engineering Department in this regard should not be overlooked. Engineering, as much as any other area, is interested in product quality in the sense of conformance to the specifications they have established, because such conformance gives the product its best chance of serving the market and justifying the design. Furthermore, Engineering often knows and can demonstrate the "why's" of some of the specifications that don't appear logical to Manufacturing or other personnel. Quality people will find Product Engineering a natural and helpful ally in their educational efforts.

Of course, the evaluation of rejected and questionable material by Engineering prior to its release for use or for salvage is a familiar facet of cooperation by Product Engineering in the quality picture. The theory of the assignment is that the designer, being the person best qualified to assess the functional requirements of a part, is the person who should rewrite the specifications, in effect, for sub-standard parts to be used, or should design a modification to these sub-standard parts to make them useable. This apparently is a pretty good theory because it has been followed in industry with some success for many years.

Another way in which Engineering contributes to the quality picture is through re-analysis and revision of the original specifications based on product performance or shop experience. Many people would say that this is just a polite way of describing the correction of mistakes made previously. And that would be right! But perhaps, it is proper to be polite in this case! As an "ex" engineer, I can assure you that it is much easier to see these mistakes after they are in production than at the time when they're just lines on a big sheet of white paper. So, it is a much more creative approach, for the Quality Control people who see the end result both in product performance and in manufacturing problems, to solicit a re-investigation and possible change by Engineering than to be overly critical of the error itself. When you go to do that job, take along your statistical data, well organized and presented with a minimum of mathematical hokus-pokus and a maximum of graphics. I think you'll find the engineer most receptive.

I am employed by the Automotive Division of the A. O. Smith Corporation and our product is chassis frames and wheel suspension parts for the automobile industry. These parts are assemblies of relatively heavy gage stampings and, while they may seem rather crude in comparison with many other automotive and industrial components, they are not without quality problems. This is partly because of the inherent characteristics of large, rangy stampings and partly because of the cost-performance-appearance equation in the automobile industry itself. I said earlier that one can hardly assess quality performance without knowing something of cost. Certainly this is true in the automotive

industry and the methods practical in our market don't allow much leeway for premium cost operations. The parts are made and assembled by press work and welding and the resulting tolerances are those of press work and welding. We feel that we have made some pretty significant quality improvements in the past several years and these may serve as examples for our discussion here. But in order to apply these examples, you will have to visualize our particular type of manufacture in which low cost is a cardinal virtue and dimensional accuracy sufficient to meet requirements must be accomplished in spite of springback, distortion following press operations, and dimensional changes resulting from the heat of welding.

In the past four to six years, there has been a healthy tendency in our industry for customer and supplier engineers and Quality Control people to meet together in advance of each series of new models and establish agreements on realistic tolerances and accuracy requirements. Prior to this period, it often happened that the Engineering Departments would apply tolerances appropriate for machined parts but not appropriate for pressed steel. Inspection Departments at both the supplier's and the customer's plants would disregard these drawing tolerances and base their acceptance on what they figured would work.

A specific example of such a condition is shown in Figure #1 which compares the drawing specification for a hole location tolerance in a truck side member with actual shop performance. The thing that prompted the study reflected here was a modest amount of assembly trouble which developed due to hole location variations. But the difference between specified and actual performance had existed for many years without previous problem so it is obvious that the slight amount of trouble being experienced was not consistent with the appearance of this chart. A joint effort on the part of Engineering and Quality Control both at our own and our customer's locations resulted in a realignment of the specifications consistent with actual requirements and an improvement in the hole location accuracy as shown in Chart #2. In this case, a 25% reduction in the total spread was accomplished and gave satisfactory performance results. The original specification would have required an 85% improvement in performance and unthinkable cost penalties. The performance improvement was achieved by taking advantage, from an engineering standpoint, of facts developed in a statistical analysis by the Quality Control Department. Quick measuring methods were developed permitting tests to be run which established the distribution of the total error among the various productive operations involved. Analysis of these variations allowed the selection of those operations capable of being improved and the calculation of the expected resultant overall gain. Engineering tests were made to substantiate the adequacy of the expected improved product and, these tests being successful, processing changes were released to suit. Production results verified the statistical analyses and service performance confirmed the tests with the result that, so far at least, we have been living happily ever after.

Often the Product Engineering group suggests, or helps to develop new processing methods which are a necessary part of a new product design. In this way having gone through a program such as that described above, we find a new model can be approached with a fair degree

of logic. We now have statistical data on the history of conformance on each of the dimensions listed. Customers' Quality Control can report on any problems they experienced with our material during the past year and this can be correlated with our performance against known specifications. At the same time, Engineering can explain the features of the new design which require special consideration for proper functioning. The principle that tolerances, like speed laws, should be as unrestrictive as possible but should be rigidly enforced, is becoming well accepted with the result that non-essential tolerances which are difficult to hold are opened up while fancier tooling or methods are planned for critical areas in the design, based on previous history, or for areas which are required by the new model to perform some improved function. I think we all feel that we are getting more for our money on this basis.

This change in philosophy hasn't occurred all at once, but it has occurred rapidly enough to have required a fair amount of missionary work with our Line Inspection and Manufacturing organizations in the shop. I'm sure the whole subject has been complicated by the fact that it has been occurring during a period of increased complexity in the design and manufacturing requirements for our product and increased necessity for accuracy in our product as a result of design changes within the vehicle.

A more difficult problem than that of dimensional requirements has been the formalizing of visual quality standards. A lot of welding is done on our frames so weld quality gets to be a matter of contention. This must be evaluated visually on the bulk of the product, although proof tests of one sort or another may be made to validate the designs or check performance of production parts on a very limited sampling basis. This type of problem, too, has responded to cooperative effort between Product Engineering and Quality Control with the lead being taken by Quality Control and with technical assistance and even some of the personnel being supplied by Engineering. Figure #3 shows an example of a visual standard established experimentally and by independent judgment of a number of representatives of Inspection, Manufacturing, and Engineering. You will note that there are three classes of welds listed and this bit of philosophy may require some explanation. The philosophy is that only one quality level is desired--the best. It is recognized, however, that not all welds on our product are subject to the same degree of loading or the amount of importance from a failure standpoint. Therefore, Product Engineering has analyzed the importance of each weld and classified it and it has been agreed to accept, without repair, lesser degrees of quality as shown in this exhibit for the Class "B" and "C" welds.

The classification and demonstration of requirements was a great help to maintenance of adequate welding quality. Of itself, however, it was not sufficient, and the addition of our Process Control Department to the Quality Control area merits discussion here. We have found this activity most helpful during the ten months it has been operating and we believe that it typifies an approach which is fundamental to today's quality requirement.

Our Process Control group consists of a number of trained

Engineering specialists in various processing fields. There are experts in Metallurgy, and the like, including a technically trained welding specialist. The entire group is under the supervision of an engineer whose original experience was in our Product Engineering activity. The background and training of these people permits them to take a methodical, scientific approach to such problems as welding quality. By review with Product Engineering, they have determined not only which welds are critical but which portions of each of these welds are likely to initiate service difficulties. Through knowledge of welding techniques, they have established optimum procedures for each weld, paying particular attention to the critical welds and the critical weld areas so that the procedure favors quality in those particular spots.

Detail instructions have been prepared for the welders, and where necessary, changes in fixturing, electrode specification, and weld machine settings have been arranged for so that all detail requirements for optimum performance have been defined. Because this program has been established on the basis of study and fact rather than opinion, it has received the enthusiastic support of all parts of our organization. Manufacturing management has been particularly progressive in taking advantage of the information presented and gains in welding quality have been apparent in some cases on a daily and almost on an hourly basis. We think this marriage of Engineering and Quality Control is a constructive thing which will lead to improvements in many areas. Welding quality is just one example.

This paper has been directed at Quality Control people although it has dealt with the Product Engineering function. It has described the ways in which Product Engineering is a part of the overall quality picture by the very definition of its assignment, and the ways in which it can be made a part of the quality picture with a little encouragement from the Quality Control people and from Engineering management and Central Management to help out. These ways are the establishment of quality oriented dimensional, visual, and material characteristics; the development of performance tests where required for quality assurance; cooperation in the development of processing methods and the establishment of controls for these methods; the training of some of the personnel who will ultimately be engaged in Quality Control work; cooperation with Quality Control Department in educating other groups in the organization to the needs for and means of accomplishing satisfactory quality; the evaluation of questionable material and establishment of salvage procedures; and the reanalysis and improvement of the original engineering specifications. We feel that members of the Quality Control organization can do much to improve the effectiveness of Engineering in the Quality program by encouraging them to be helpful in each of the foregoing areas. Mr. Norman Shidle of the SAE JOURNAL recently quoted Dwight Morrow that "we tend to judge ourselves by our motives and others by their actions". I think this is pretty true and that Quality Control people, at least some of those with whom I have been acquainted, have not fully recognized the parallelism of aims between their own group and the Product Engineering group because they have mistaken the engineer's preoccupation with other problems as a lack of interest in the quality problem. I think you will find that the engineer is intensely sympathetic with quality requirements and you'll need only to encourage him a little bit to make him a very valuable ally. This alliance will prove beneficial to your company, to the Engineering Department, and perhaps to you yourself.

FIG.-8

CURRENT VERSUS IMPROVED HOLE LOCATION CAPABILITIES ON TRUCK PARTS

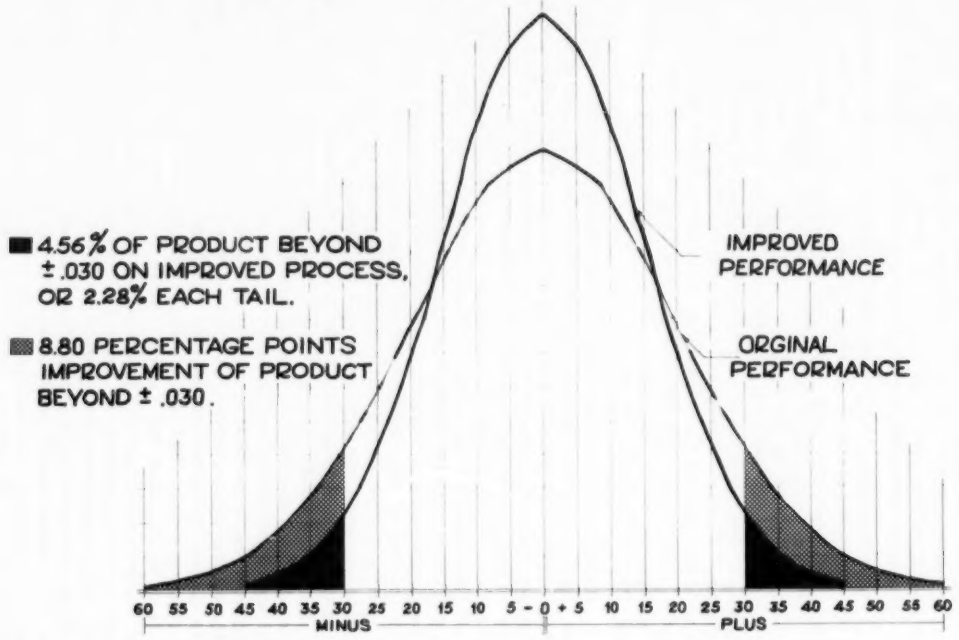
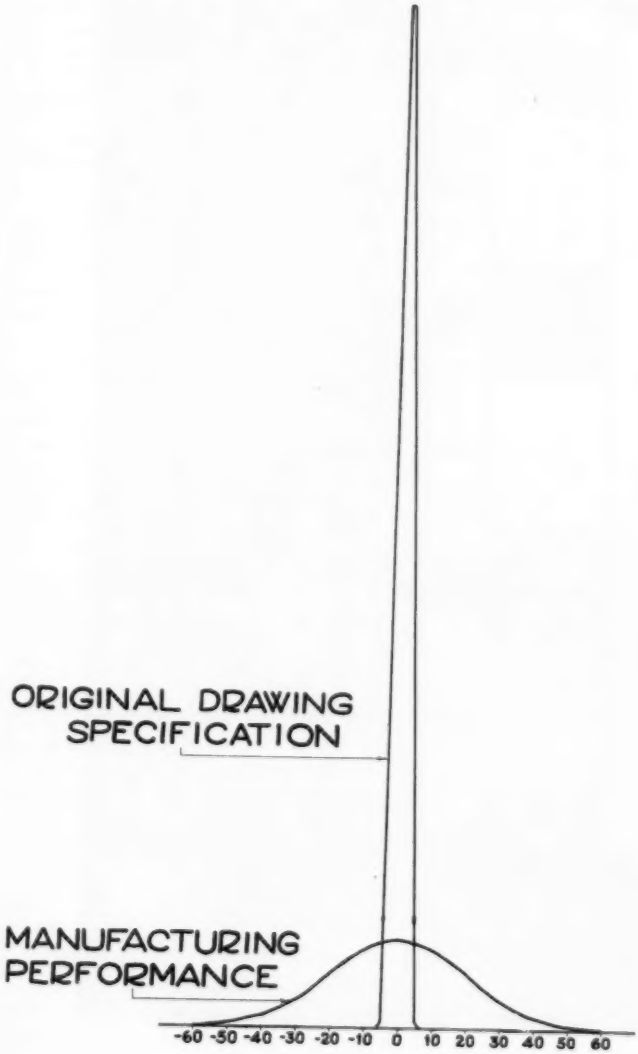
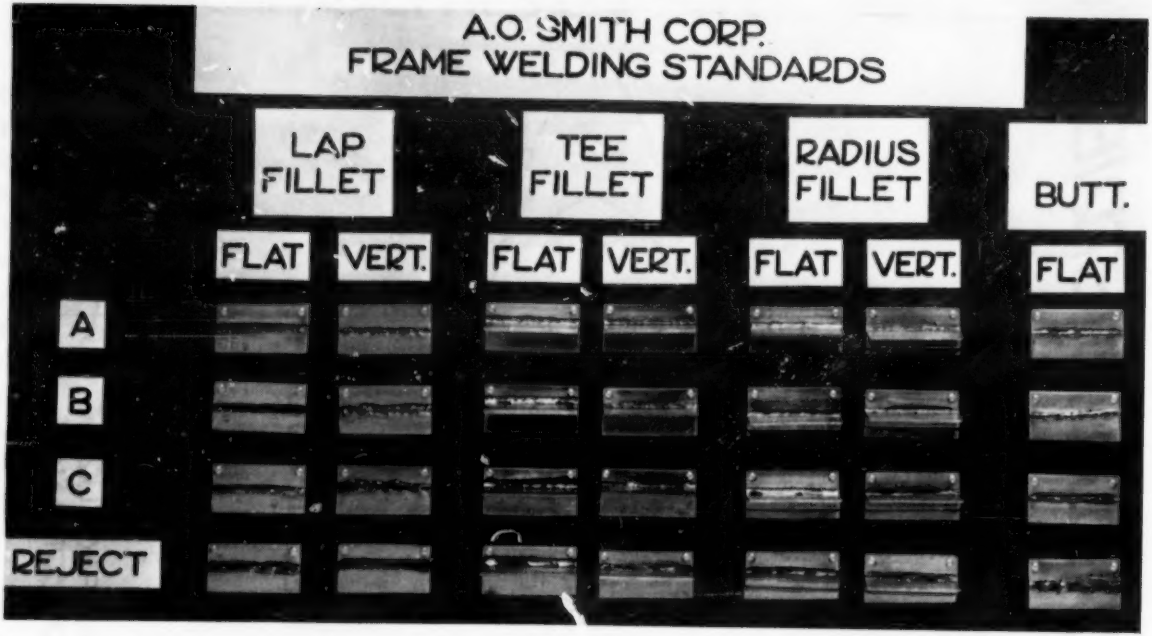

4.56% OF PRODUCT BEYOND ± .030 ON IMPROVED PROCESS, OR 2.28% EACH TAIL.
8.80 PERCENTAGE POINTS IMPROVEMENT OF PRODUCT BEYOND ± .030.
IMPROVED PERFORMANCE
ORGINAL PERFORMANCE
60 55 50 45 40 35 30 25 20 15 10 5 - 0 + 5 10 15 20 25 30 35 40 45 50 55 60
MINUS
PLUS

FIG-2

QUALITY DISTRIBUTION




A.O. SMITH CORP.
FRAME WELDING STANDARDS
LAP FILLET
TEE FILLET
RADIUS FILLET
BUTT.
FLAT
VERT.
FLAT
VERT.
FLAT
VERT.
FLAT
A
B
C
REJECT

EARLY DETECTION OF RELATIVELY SMALL SHIFTS IN A PROCESS AVERAGE

S. W. Roberts
Bell Telephone Laboratories, Incorporated

An $\bar{X}$ control chart presents a graphical summary of periodically collected information concerning a process average. Such information is useful in making decisions about the presence or absence of trouble in the process. These decisions are subject to two types of errors: (1) Type 1 errors are erroneous decisions that trouble is currently present, and (2) Type 2 errors are erroneous decisions that no trouble is present. A Type 1 error results in wasted effort of looking for non-existent trouble, while a Type 2 error allows the trouble to remain undetected for one extra period of time.

Many different tests may be employed in making decisions concerning the presence or absence of trouble. The standard control chart test[1] indicates that trouble is present if a single point falls outside the 3σ limits. Shewhart[2] noted that in many applications such a test seemed to strike a balance between the economic consequences of Type 1 and Type 2 decision errors.

The spread of the use of control charts has been accompanied by a growth in the number of applications where the standard control chart test is either replaced or supplemented by other tests. In many applications 2σ limits are used rather than 3σ limits; in others, action is taken when certain types of runs occur - for example, when seven consecutive points fall on one side of the central line of the control chart. Numerous other tests are also in use or have been proposed. (References 3-10.)

To facilitate comparisons of several types of tests, we assume that we have a process that is in control with a known process average μ_o and standard deviation σ' except for periods following the occurrence of trouble, which shifts the process average to a new level $\mu_o + \delta$ where it remains until detected. As a point $\bar{X}_j$, representing the average of n measurements taken at time j (j = 1,2,3,...), is plotted on the control chart a decision is made concerning the hypothesis that no trouble is currently present, or, equivalently, that the current process average $\mu = \mu_o$, or that $\delta = 0$. The effectiveness of a test is summarized by a curve of E(y) versus Δ, where E(y) is the expected number of points from the time the trouble occurs until it is detected, and $\Delta = \delta/\sigma_{\bar{X}}$ is the amount of the shift measured in units of $\sigma_{\bar{X}} = \sigma'/\sqrt{n}$. We present curves of E(y) versus Δ for some run tests and for tests based on two types of moving averages.

Figure 1 presents a set of curves showing the effect of changing limit lines. These curves are for tests $T_1(L)$ that reject the null hypothesis if a single point falls outside the limit line at $\mu_o + L\sigma_{\bar{X}}$. We shall consider one-sided tests in this paper; for most practical purposes we can obtain sufficiently accurate curves for the corresponding two-sided tests by taking half the ordinate at $\Delta = 0$ and drawing a smooth curve, with zero slope at $\Delta = 0$, to approach the curve for the one-sided test in the neighborhood of $\Delta = 1/2$.

The ordinate of a curve at $\Delta = 0$ has a special meaning - it represents the expected number of points between consecutive Type 1 errors when the process remains trouble-free. For $\Delta > 0$, $E(y) - 1$ is the expected number of consecutive Type 2 errors made following a shift of magnitude Δ .

Comparing the curves of Figure 1 for tests using 3σ limits and 2σ limits, we see that the latter test detects all shifts sooner, on the average, but the cost is in terms of more frequent Type 1 errors; Type 1 errors are over sixteen (741/44) times as frequent, while the expected number of consecutive Type 2 errors is reduced for all values of $\Delta > 0$, for example, from 43 to 5.3 at $\Delta = 1$. The choice between the use of 2σ or 3σ limits in a particular application should be based on the relative consequences of Type 1 and Type 2 errors, since the tests are equally simple to perform.

Any desired average frequency of Type 1 errors can be effected by an appropriate setting of the limit lines. The curves of Figure 1 are horizontal translations of one another; the amount of the translation corresponds to the amount by which their limit lines are translated from one another on the control chart. Thus the curve for $T_1(2)$ is obtained by translating the curve for $T_1(3)$ one unit to the left. This procedure can be followed for curves for other tests, though in general the result is only an approximation.

Figure 2 shows curves for tests $T_k(L_k)$ that reject the null hypothesis if k (k = 1,2,3,4,6,8) consecutive points fall outside limit lines at $\mu_o + L_k\sigma_{\bar{X}}$, where the L's are chosen so that all tests have the same average frequency of Type 1 errors. Notice that each test has an interval of Δ in which its curve is lower than the others. In particular, the standard control chart test detects relatively large changes sooner than the other tests, on the average.

Figure 3 illustrates the result of combining two tests - the curve for the resultant test is lower than those for the individual tests. Here we consider a test $T_{1,7}(3,0)$ that rejects the null hypothesis if a single point falls outside the 3σ limit or if seven consecutive points fall above the central line of the control chart. Every additional criterion for rejecting the null hypothesis lowers the curve, indicating that changes are detected sooner at the expense of more frequent Type 1 errors. In addition, the resultant test becomes more complicated to perform as additional criteria are considered. With this last fact in mind, there is a question of whether we might not do almost as well by using a test based on a single point, with the limit line selected so that the test has the same expected frequency of Type 1 errors as the more complicated test. In our example of Figure 3 the dashed line applies for the test that rejects the null hypothesis if a single point falls outside a limit line at $\mu_o + 2.56\ \sigma_{\bar{X}}$. The dashed line crosses the curve for the more complicated test $T_{1,7}(3,0)$ at about $\Delta = 1.5$; for applications where changes represented by $\Delta > 1.5$ are of primary concern it evidently would not pay to use the more complicated test (for sample size n = 4, $\Delta = 1.5$ corresponds to a shift of $.75\sigma'$).

In general, run tests do not make efficient use of the information supplied by the control chart points. In this respect tests based on moving averages of consecutive points are better, as illustrated by

Figure 4, which pertains to tests $M_k(L)$ (for Figure 4, L = 3) based on the moving averages of k (k = 1,2,4,8) consecutive control chart points, with limit lines at $\mu_o + L\sigma_{\bar{X}}/\sqrt{k}$. A possible disadvantage of moving average tests is that they may not be as simple operationally as run tests. However, in some cases it may be relatively easy to compute such averages, and in other cases they can be generated graphically, based on the fact that the point halfway between two control chart points represents their average value.[8] A test based on the moving average of two consecutive points would be particularly simple to perform.

A simple graphical procedure[10] can be used to generate points representing a moving average in which the entire history of points is assigned weights, with the weights decreasing as a geometric progression from the most recent point back to the first point. We shall refer to such moving averages as geometric moving averages. The current geometric moving average, Z_j, is defined by the equation

$$Z_j = (1-r)Z_{j-1} + r\bar{X}_j, \quad j > 0, \quad 0 < r \leq 1,$$

which indicates that the point Z_j is located a fraction r of the distance from Z_{j-1} to $\bar{X}_j$ on their connecting straight line.

To generate the Z's graphically, the following procedure is suggested:

(1) Mark $\bar{X}_1$ at abscissa 1.

(2) Mark $Z_0 = 0$ at abscissa $1 - 1/r$.

(3) Draw a straight line connecting Z_0 with $\bar{X}_1$; the point on this line with abscissa $1 - (1-r)/r$ is Z_1.

The next point, Z_2, will be at abscissa $2 - (1-r)/r$ on the straight line connecting Z_1 with $\bar{X}_2$. Both the X's and the Z's progress one abscissa unit at a time, with Z_j always $(1-r)/r$ units to the left of $\bar{X}_j$. The standard deviation of Z_j approaches a value of $\sqrt{r/(2-r)}\ \sigma_{\bar{X}}$ as j increases.

Figure 5 shows curves for tests $G_r(L)$ (for Figure 5, L = 3) that reject the null hypothesis when Z_j falls outside limit lines at $\mu_o + L\sqrt{r/(2-r)}\ \sigma_{\bar{X}}$, for r = 1/4, 1/2, and 1. An r of 2/5 is suggested as an experimental value.

If points representing moving averages are to be generated on the control chart, it might be of some practical, if not theoretical, importance to have limit lines for the $\bar{X}$'s as well as for the moving averages. Figures 6 and 7 present curves for tests $M_{1,k}(3,3)$ and $G_{1,r}(3,3)$, respectively; these tests may be considered to be the standard control chart test supplemented by the tests of Figures 4 and 5 involving ordinary moving averages and geometric moving averages, respectively.

Because the ordinate scales are logarithmic in Figures 1-7, it is difficult to appreciate the savings in E(y) for the various tests as compared to the standard control chart test. Figure 8 shows such savings on an arithmetic scale. The top curve applies to particular

moving average tests (M_8 (2.95) or $G_{1/4}$ (2.95)), and the lower curve applies to run test $T_{1,10}$ (3.13,0) that rejects the null hypothesis if (1) a single $\bar{X}$ falls outside a limit line at $\mu_o + 3.13\,\sigma_{\bar{X}}$, or (2) ten consecutive $\bar{X}$'s fall above the central line of the control chart. The curves go through the origin because the limit lines were chosen to provide a common level of Type 1 errors.

A comparison of the various curves presented, and a consideration of the relative simplicity of the testing procedures, indicates that for relatively large changes tests such as the standard control test that base their decisions on a single control chart point are most practical. By relatively large we mean changes greater than $2/\sqrt{n}$ or $3/\sqrt{n}$ units of σ'. For relatively small changes, other tests of the types considered here may prove useful. In some cases it may be necessary to increase the sample size or frequency of sampling; these two alternatives have been largely eliminated from consideration in this paper, for we are essentially trying to make efficient use of the given information supplied by a history of control chart points.

References:

1. American War Standard, "Control Chart Method of Analyzing Data", American Standards Association, Inc., New York, N. Y., Z1.2-1941.
2. Shewhart, W. A., Economic Control of Quality of Manufactured Product, D. Von Nostrand Co., Inc., New York, N.Y., 1931.
3. Grant, E. L., Statistical Quality Control, McGraw-Hill, New York, 1946.
4. Weiler, H., "The Use of Runs to Control the Mean in Quality Control", Journal of the American Statistical Association, Vol. 48, pp. 816-825, 1953.
5. Page, E. S., "Control Charts with Warning Lines", Biometrika, Vol. 42, pp. 243-257, 1955.
6. Statistical Quality Control Handbook, Western Electric Co., New York, 1956.
7. Kitagawa, T., and Seguchi, T., "The Combined Use of Runs in Statistical Quality Control", Bulletin of Mathematical Statistics, (Japan), pp. 24-45, 53-72, Dec. 1956, Mar. 1957.
8. Roberts, S. W., "Properties of Control Chart Zone Tests", Bell System Technical Journal, pp. 83-114, January, 1958. Bell System Monograph 2938.
9. Toda, H., "Band-Score Control Charts", Reports of Stat. Applications Research, Union of Japanese Scientists and Engineers, Vol. 5, No. 2, pp. 20-24, 1958.
10. Roberts, S. W., "Control Chart Tests Based on Geometric Moving Averages." Submitted to Technometrics for publication.

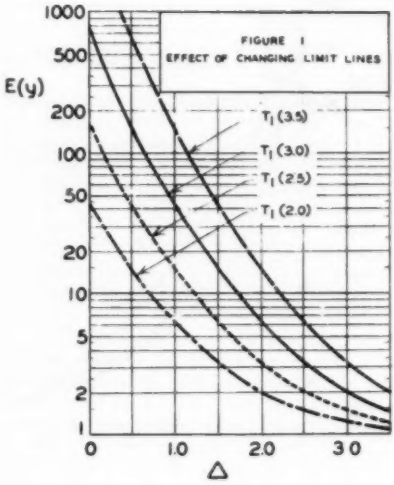


FIGURE 1
EFFECT OF CHANGING LIMIT LINES
E(y)
1000
500
200
100
50
20
10
5
2
1
0
1.0
2.0
3.0
Δ
T1 (3.5)
T1 (3.0)
T1 (2.5)
T1 (2.0)

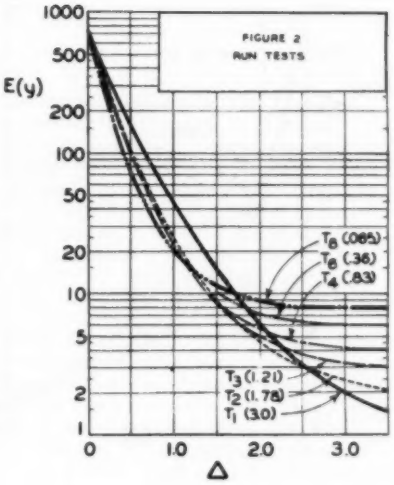


FIGURE 2
RUN TESTS
E(y)
1000
500
200
100
50
20
10
5
2
1
0
1.0
2.0
3.0
Δ
T6 (.065)
T6 (.36)
T4 (.83)
T3 (1.21)
T2 (1.78)
T1 (3.0)

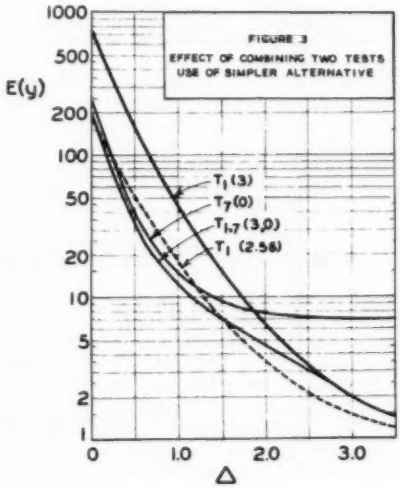


FIGURE 3
EFFECT OF COMBINING TWO TESTS
USE OF SIMPLER ALTERNATIVE
E(y)
1000
500
200
100
50
20
10
5
2
1
0
1.0
2.0
3.0
Δ
T1 (3)
T7 (0)
T1,7 (3,0)
T1 (2.56)

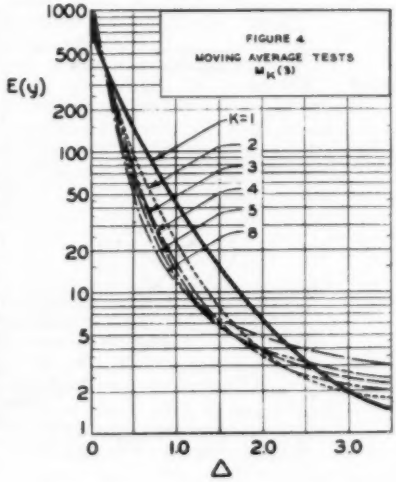


FIGURE 4
MOVING AVERAGE TESTS
MK(3)
E(y)
1000
500
200
100
50
20
10
5
2
1
0
1.0
2.0
3.0
Δ
K=1
2
3
4
5
8

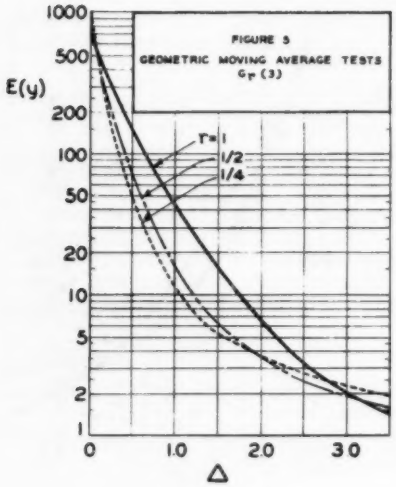


FIGURE 5
GEOMETRIC MOVING AVERAGE TESTS
$G_\Gamma(3)$
E(y)
1000
500
200
100
50
20
10
5
2
1
Γ=1
1/2
1/4
0
1.0
2.0
3.0
Δ

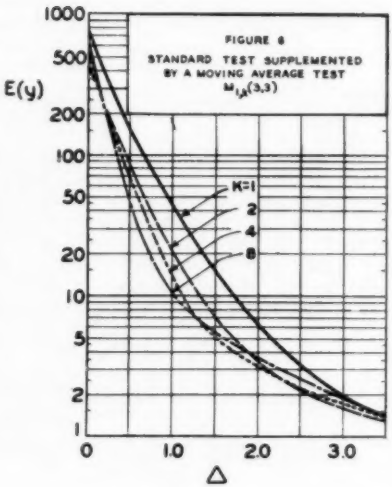


FIGURE 6
STANDARD TEST SUPPLEMENTED
BY A MOVING AVERAGE TEST
$M_{1,K}(3,3)$
E(y)
1000
500
200
100
50
20
10
5
2
1
K=1
2
4
8
0
1.0
2.0
3.0
Δ

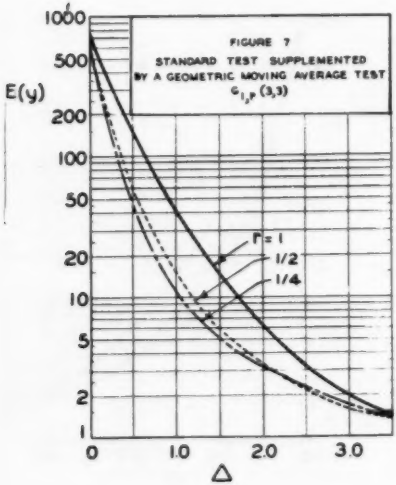


FIGURE 7
STANDARD TEST SUPPLEMENTED
BY A GEOMETRIC MOVING AVERAGE TEST
$G_{1,\Gamma}(3,3)$
E(y)
1000
500
200
100
50
20
10
5
2
1
Γ=1
1/2
1/4
0
1.0
2.0
3.0
Δ

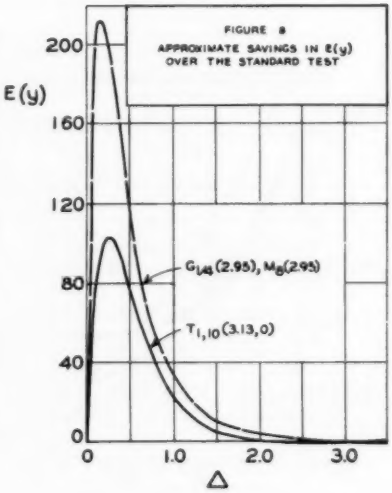


FIGURE 8
APPROXIMATE SAVINGS IN E(y)
OVER THE STANDARD TEST
E(y)
200
160
120
80
40
0
$G_{1/8}(2.95)$, $M_8(2.95)$
$T_{1,10}(3.13,0)$
0
1.0
2.0
3.0
Δ

SOME FOUNDRY APPLICATIONS OF STATISTICAL QUALITY CONTROL

Sigmund P. Zobel
The Carborundum Company

INTRODUCTION

It is the purpose of this paper to discuss briefly a few examples of applications of statistical quality control principles in a steel foundry. Space limitations preclude the discussion of a total system for quality control in the foundry, so only a few of the potential areas or usages will be discussed. The reader will find little or nothing new if he is a student or practitioner of statistical quality control, and is seeking new techniques or methodology. Quality control problems to be coped with in a foundry are really no different from quality control problems to be encountered in other industries. Relative emphasis may differ. That is all.

In looking for the quality-sensitive areas of the foundry operation, one is led to the same as in any other situation - the raw materials, the processing, the fabrication, the machines and equipment, the materials handling, the operators, gage and instrumentation control, etc. One outstanding characteristic that may differ in the foundry compared to many other places is that it is frequently a job shop, where some of the production runs may be as short as one piece. Contrary to popular misconceptions, this does not preclude the successful application of SQC. By placing the emphasis on processes, functions, or even personnel, we get the conditions of continuity and homogeneity requisite for the application of probability-founded principles.

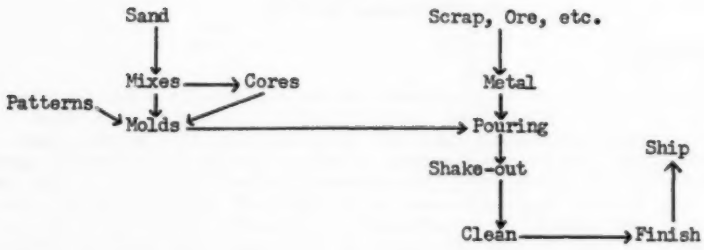
In the foundry, as in any other shop, it is well to begin the planning for statistical quality control with a flow sheet for the purpose of constructing a model of the processing. From the flow sheet it is possible to identify where operations are performed (or transports or holds, etc.) which can affect the quality of the finished product. An abridged flow sheet, without the use of conventional symbols, for a sand foundry appears as Illustration 1.

WHERE SQC CAN BE USED

The form of the flow sheet in Illustration 1 was deliberately chosen to delineate the most useful starting points for installing or expanding a reasonably comprehensive system for control. Each item named, be it indicative of raw material or actual processing, can be subjected to one or more statistical controls useful in building quality into the castings at reasonable costs. This paper will discuss applications made to sand and mix control, molding machine control, melt control, finishing control, and one or two administrative applications in areas of cost control. These applications have been selected for the present purpose since they are probably the easiest to implement and can do the most good in the shortest time.

Illustration 1

Basic Flow Sheet for
a Sand Foundry



SAND AND MIX CONTROL

At least two procedures are useful as control devices for the sand going into the mixes. Both have to do with the fineness distribution of the grains of sand. In the first instance, a procedure for making decisions on acceptance or rejection of carloads of new sand was devised. On the basis of the records of samples from the last fifty acceptable shipments, a standard distribution of per cent retained by sieve size was derived. After testing on normal probability paper for validity of the standard so derived, upper and lower specifications were drawn up for each sieve, indicating the highest and lowest percentages tolerable according to the sand technicians. From these, a quasi-control chart for acceptance decisions could be devised. Figure 1 shows the standard and specification values, while Illustration 2 shows the chart itself. With the aid of the chart depicted in Illustration 2, it was no longer necessary to deliberate on the results of the samples from each carload before acceptance or rejection; the decision became automatic, except for borderline cases.

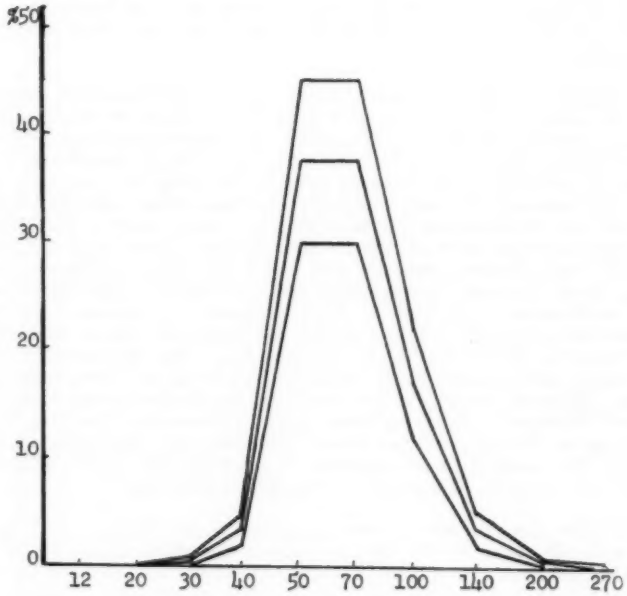
Figure 1

Standard For New Sand Sampling

Sieve Size	Standard % Retained	Lower Limit	Upper Limit
30	0.5	0.0	1.0
40	3.5	2.0	5.0
50	37.5	30.0	45.0
70	37.5	30.0	45.0
100	17.0	12.0	22.0
140	3.5	2.0	5.0
200	0.5	0.25	0.75
270	0.0	0.0	0.50

Illustration 2

New Sand Acceptance Sampling Chart



The second application was on a similar sampling procedure, but was more like a process control. Needless to say, reclaimed sand was used in many mixes. If the sand scrubber broke down or for any other reason failed to yield sand of the desired properties, production delays might result. What was needed was a sampling procedure and analysis that would facilitate detection of impending trouble or indicate its existence as soon as it had occurred. Illustration 3 shows the solution that satisfied the objectives. It is a stratified graph showing what part of the grain distribution is contributed by each sieve. (Although for clarity of illustration, not each sieve is shown separately.) So long as the bands remain fairly constant in vertical width between boundaries, the grain fineness distribution may be assumed to be stable. A malfunctioning of the system was revealed by the high peak about midway through the graph.

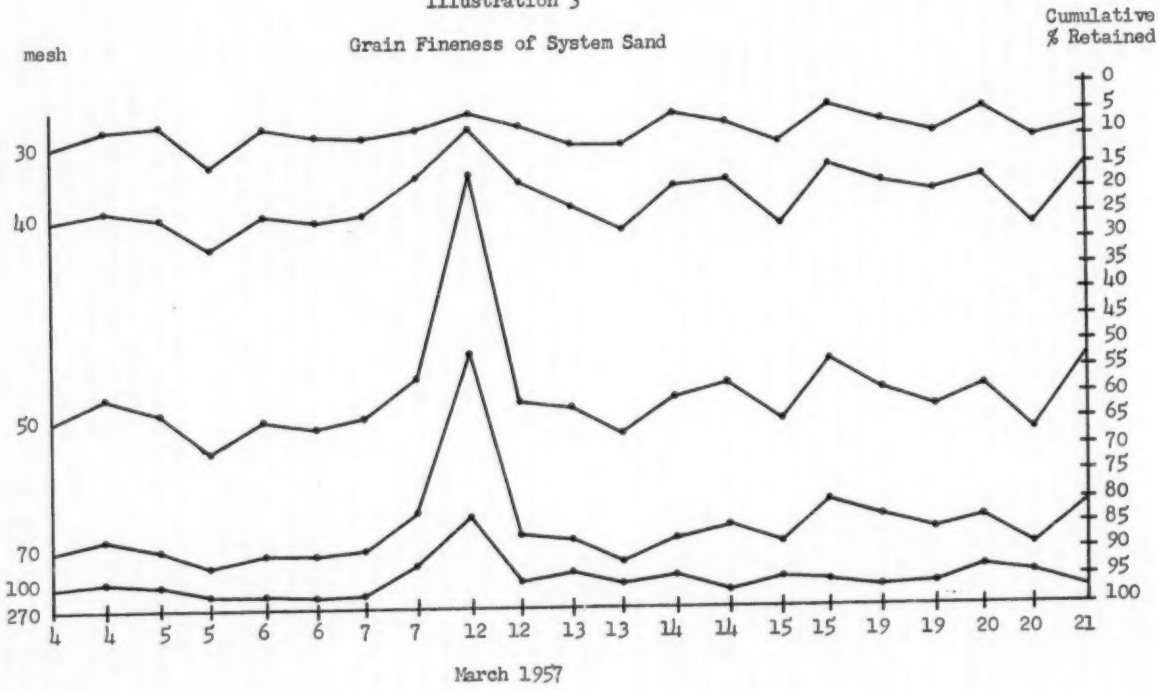
Turning to the mixed sand, there are certain physical properties which distinguish a good batch of sand from a poor batch. These properties are denoted by specifications on such characteristics as moisture, permeability, compression, deformation, etc. Even though there may be scores of different jobs running through the shop in a week or month or year, the chances are there are only a couple dozen sand mixes that are made for all the different jobs. And, of these, only a small number may be highly critical with regard to the quality of the casting. For example, this foundry operated with about two dozen or fewer standard mixes, which included facing sands and core sands. The finish of a steel casting was important to the customers. To maintain and/or improve this quality characteristic, averages and ranges charts were instituted on the principal facing sand mix. The limits were not empirically determined, as is usually the case, but were obtained by working backward from the specifications on the two properties of greatest interest. This was done in the interests of keeping the objective in front of the personnel from the very beginning of the program, and because it was desired to provide the psychological stimulation of a form of challenge to the mill operators.

To give the procedure every opportunity to become a real control, a board was installed on the railing around the sand muller where this mix was made, in such a position that the operators had to pass it frequently. The tests made from four successive batches comprised one sample, and the results plotted as they became available, with different lines for each of the two shifts maintained on that muller.

Initially, there was considerable variation from batch to batch, and the levels, as indicated by the averages, were not satisfactory. However, the operators showed considerable interest in the points and lines and frequently asked what they meant to make sure they understood. In a matter of days, the ranges were under excellent control, and the averages were under good control, mainly due to the interest and cooperation of the muller operators with the sand technicians. As a result of the success here, the sand technician requested instituting a similar control for one of his newer core sands, before he found difficulty. The interesting thing in this request is that only five months earlier this same man was very skeptical as to what SQC could accomplish in his area.

Illustration 3

Grain Fineness of System Sand



MOLDING CONTROL

The daily scrap was analyzed in such a way as to indicate which defects may be attributed to molding practice, which to pouring practice, finishing, etc. The scrap due to molding practice was charged back to the machine on which it was molded. Scrap rates were published weekly on a p chart for each machine, as well as for all molding machines combined, with the usual benefits of seeing trends or unusual deviations from past performance.

On a daily basis, principal defects, whether they caused scrap or not, were charted on the wall of the foundry office so that they may be closely observed by the supervision most concerned. Since most defects may be attributed to a specific phase of the foundry operation, this was invaluable for helping the Foundry Superintendent spot trouble areas. As a further aid in this direction, and along with the weekly scrap by machine rates, weekly scrap rates for each operating department were published and all the scrap combined into one weekly plant-wide scrap rate, which was also put on a p chart. The simple expedient of putting an upper limit on the plant scrap rate, and sending a special "quality deviation" form whenever a point exceeded the control limit or there was a significantly long run up or down got department heads to being sufficiently scrap-conscious so that in three month's time the rate was cut in half and subsequently to one third the former level.

OTHER AREAS OF CONTROL

Next, we turn our attention to the melting department, once again necessarily limiting the discussion to just one or two applications.

For each of six important elements, the % within the specifications in the last twenty heats were plotted against its respective limit. When any point was out of control, or an undesirable trend was observed, the melting superintendent would be informed of the condition. In addition, a similar chart was maintained which gave the % of heats of the last twenty which were out of spec for any element. To supplement this, what was called the melters' performance chart was maintained. This showed the proportion of heats within specifications for each melter not only for the basic steel made but for any others. Weekly, the melting superintendent was given the melters' performance charts for his information. And you can be sure that the melters themselves watched their charts very closely, to make sure they were not done an injustice, as well as to see how they were doing.

Now, consider briefly how SQC may be applied to the finishing operations. Here, as elsewhere, there is much that can be done. But, cleaning, heat treat and finishing operations are so hectic in most foundries, getting data which are reliable enough to make use of requires more personnel than management is usually willing to permit for the job. However, much could be done from the daily reports filed by inspectors. One such application, the daily chart of principal defects found, has already been cited. But this is not directly aimed at controlling efficiency and costs in finishing itself.

A good beginning along these lines can be made by maintaining control charts on the % of rework required. For example, on one piece made in relatively large volume, the foreman concerned was provided with a daily p chart on each of several reasons for reject which required rework, such as bore not properly ground, or walls not properly ground, etc. This information, which came from the daily report of an inspector whose station was at the end of the grinding line, proved to be indispensable in maintaining a high rate of quality production on this piece. Reports on customer returns may also be analyzed to shed light on the effectiveness of the finish and heat treat operations. In addition, they are useful in evaluating the effectiveness of the final inspection.

The applications presented thus far have almost all been concerned with processes. The fact that a foundry is essentially a job shop has not been of any hindrance. And this is a point which cannot be emphasized enough. So long as a process is relatively stable, control of the end products can be maintained even though the product mix itself changes.

The process approach may be underscored by discussing briefly the average job concept. This is a means of scrap control devised for a foundry, or almost any job shop, and appeared in the March 1957 issue of Foundry, under the title of "The Average-Job Quality Control System," by Kenneth Smith.

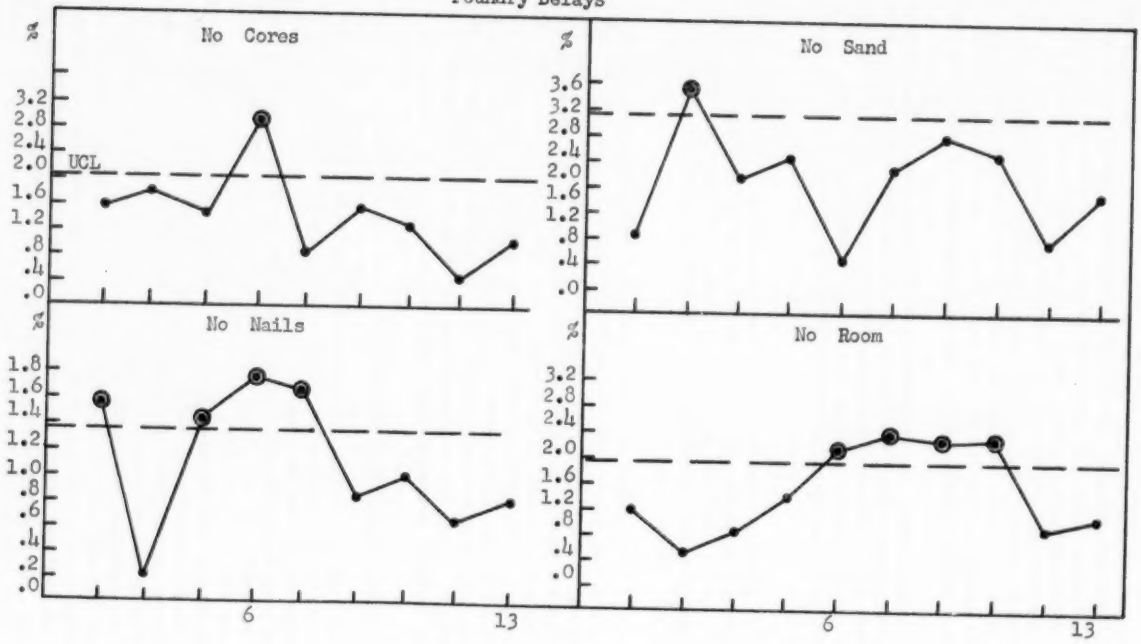
The total number of pieces made for all patterns in some base period and the number of pieces scrapped may be used to obtain an average scrap rate, or $\bar{p}$. This scrap rate is considered to be the average performance level for each pattern. For each day, or week, or whatever time period you wish to work with, the scrap rate for each pattern is determined. The rate of scrap for any pattern is not considered excessive unless it exceeds its particular upper control limit, as found by using the $\bar{p}$ previously found and the number of pieces inspected as the sample size. An approach such as this greatly broadens the spectrum of statistical controls in a job shop.

ADMINISTRATIVE AREAS OF APPLICATION

All of the foregoing applications have dealt with various aspects of production itself. But there are other areas in the foundry where SQC procedures can be helpful in the interests of cost control or reduction, which is, of course, the primary objective of all quality control work. As an example, for one reason or another, the productive workers may be prevented from doing their normal routines. This time is typically called delay time. A certain amount of this is unavoidable, but proper supervision and preventive maintenance can minimize the time and hence money spent on delays. Considerable success was realized in decreasing delay time in several areas by the use of p charts. To give one such example, daily charts were maintained on the "controllable delay" per cent of total man-hours for core-making, core-pasting and molding, by related groups of delays. When any point exceeded the upper control limit, foundry or core room supervision was asked to provide, in writing, an assignable cause, if one can be determined. This kept everybody from getting excited over delays that were really not unusual, and the underlying causes of excessive delays often fell into a discernible

Illustration 4

Foundry Delays



December 1957

pattern, on the basis of which corrective action could be taken. A sample of these charts appears as Illustration 4.

Similar charts, p charts, proved useful in controlling such diverse subjects as the honesty of piece-workers and down time of presses.

CONCLUSION

An attempt was made in this paper only to relate a few instances where basic SQC tools have been found helpful in a sand foundry. Many very good statistical tools for the foundryman have not been discussed in the interests of brevity. However, the statistician in the foundry, as well as virtually any other plant statistician, should be prepared to use the tools of statistical inference, testing of hypotheses, design of experiments and operations research. It is for this reason that the references cited below do not include the many papers or the one or two books treating foundry quality control, but include a small number of the general books which can be of help in implementing regular or ad hoc procedures for effective control of quality.

REFERENCES

Dixon, W. J. and Massey, F. J. Jr., Introduction to Statistical Analysis, McGraw Hill Book Company, Incorporated, Second Edition, 1957.

Duncan, Acheson J., Quality Control and Industrial Statistics, Richard D. Irwin, Incorporated, Revised Edition, 1959.

Volk, William, Applied Statistics for Engineers, McGraw Hill Book Company, Incorporated, 1958.

RELIABILITY, QUALITY CONTROL, AND SIMULATION

Albert S. Cahn
Mathematician
The RAND Corporation

Among workers in the field of reliability and quality control, one frequently encounters a line of reasoning that goes like this: reliability is obviously a good thing; a more reliable item, per se, is preferable to a less reliable one; and, consequently, almost any additional system of quality control is worth its cost if it increases reliability sufficiently. Up to a point, all this sounds logical.

I am sure we have all heard horrendous tales of submarines which, at great peril, have successfully intercepted enemy battleships, only to have all their remaining torpedos fail. We have heard of soldiers' needlessly losing their lives because machine-guns jammed; there have been bombing missions that were ineffective, because some unreliable component or weapon failed.

Industry can furnish many similar stories of expensive and elaborate undertakings which either failed or were only partially successful because of the poor reliability of some items. Naturally, such misfortunes give rise to relentless and frequently expensive improvement programs.

Usually, in such programs, it has been the habit to estimate the cost of the enterprise -- the experiment, the bomber raid, the submarine voyage, or whatever the enterprise is -- and to multiply this cost by an estimated increase in the probability of success deriving from programs aimed at increasing the reliability of the components used. This product, the cost of the enterprise multiplied by the estimated increase in probability of success, is often taken as the value of a reliability-improvement program.

I am sure that this estimated monetary value of increased reliability has often impelled various organizations to press hopefully forward in launching new and costly quality control programs, or programs to produce improved, more reliable -- albeit more expensive -- components.

With a little reflection, however, one can see that this method of evaluating the worth of a reliability program is liable to be faulty. The basic error usually lies in assuming that the new program is the only correct or optimal one in the light of the cost of achieving greater reliability.

For example, let us suppose that the cost of a reliability program in the manufacture of a certain bomb is ten million dollars. Let us suppose that this program would raise the probability of the bomb's functioning properly from 0.5 to 0.99. This is so marked an improvement that the program sounds immediately plausible. However, suppose that for this same ten million dollars we could so modify the bombing planes that they could carry three times as many bombs on a raid. Clearly, a raid will do more damage if it is possible to drop three times the number of bombs, even if they are still only half as reliable as they might be.

In other words, for proper evaluation of the worth of any program of quality control or reliability improvement, we must take account of reliability as a function of cost, it is true, but we must also consider the optimal use of any particular item as a function of its reliability. It is only by weighing both the reliability of an item and its use that the true value of reliability becomes clear. Viewed in this light, the cost of achieving increased reliability becomes only one of a number of costs in accomplishing a certain goal. It is the total cost which we wish to minimize and, oddly enough, it may be that we can do it most cheaply and effectively by using unreliable but inexpensive items.

Our discussion up to this point has now confronted us with a much more difficult problem than that of merely working out a program to increase the reliability of an item, and estimating its cost. We are now saying that to make a good decision, one must first undertake an entire series of estimates which relate achievable reliability to cost, and then examine an entire spectrum of reliability values in order to plan an optimal, or at least a successful, enterprise.

And to make the problem still more difficult, there is the question of time -- almost always a major stumbling-block in programs of any real importance. Usually, because of a lead-time requirement, it is necessary to plan an enterprise and to make substantial investments in it long before one has any actual concrete evidence on the overall reliability of the various items used. Likewise, there is usually no clear-cut relationship between reliability and cost in which a planner can place much confidence.

Faced with such a problem, or, more exactly, with such a group of problems, any decision-maker faces an extraordinarily difficult task. Certainly, it is almost impossible for him to lay out any set of rules or to outline any computations which will infallibly lead him to the right answer. In such a state of frustration, flipping a coin or gazing into a crystal ball seem to be as valid as human reason, and certainly a great deal easier.

But there are other alternatives as the title of my talk suggests, and as I myself believe, there is a powerful method for attacking problems of this sort in what is generically and somewhat ambiguously called "simulation." Unfortunately, simulation itself is so vague a procedure that it is hard to give any cut-and-dried methods for using it to investigate any particular problem, or even to give any general rules for simulation design. Actually, the side or boundary conditions which prescribe the limits or type of activity of whatever enterprise is under consideration are usually overriding in the design of a useful simulation.

Perhaps an illustrative example will be the easiest way to show the value of a simulation model in answering the type of question we are discussing. I will describe in some detail a simulation which the Logistics Systems Laboratory of The RAND Corporation is currently conducting, and then outline some general rules which may be helpful in designing and building any specific simulation model.

About a year ago the Logistics Systems Laboratory began investigating, in broad detail, the logistics implications of the Air Force's

ICBM program. This is obviously an extensive and difficult study of a many-faceted, important problem. The objective of the missile force -- to achieve a formidable and unquestionable deterrent posture -- is perhaps as important an objective as any which the Department of Defense now has. And the cost of our missile force will be a major item in the country's defense budget during the next several years. Certainly, any appreciable percentage saved as a result of preliminary study can be a large amount.

It became apparent early in our study that this was too large a problem to be solved in one stroke. For any hope of success we were forced to consider portions of the general problem, in the hope that we would later be able to piece our findings together into a coherent and intelligible picture. Thus, our first step was to simulate the operations of a missile squadron -- a military example of the kind of "enterprise" I discussed at the start.

In this case, the enterprise is the operation of the squadron, whose objectives are:

(1) To be able to fire a certain minimum number of missiles within a certain time-limit, with a required confidence level, and,

(2) To fire as many more missiles as possible upon demand.

The military planner has several variables at his disposal which he can adjust and manipulate in seeking the best procedure for accomplishing these objectives. The more important variables are:

(1) Number of missiles in the squadron.

(2) Geographic configuration of the squadron. (Number of missiles per launching site; number of launching sites; distance between sites; distance from sites to squadron headquarters.)

(3) Available resources. (Personnel, equipment, stocks of spare parts.)

(4) Degree of hardening of launch complexes. (Amount of overpressure complex is able to withstand without damage to missiles.)

(5) Reliability of missiles, equipment, spare parts.

(6) Communication and transportation facilities.

In addition to these variables which the military planner can adjust as he sees fit, there are a number of side conditions which may strongly influence his operation; and even though he may be unable to control them, he must nevertheless consider them and include them in any simulation. In the example at hand, these are such items as weather conditions, length of time required to obtain additional resources from a depot or factory, probability and nature of enemy action, etc.

In order to handle such a large complex of interrelated variables, we decided that the best approach was to start out with a model which combined man and computing machine, and hope we would be able to follow

this with an all-machine simulation. It was our idea that by starting with a man-machine model we would gain a good evaluation of the realism and completeness of the simulation, principally from discussion with the participants. Likewise, we felt that in this way we would know that the planned information flow of the simulation was adequate for proper operation, and would thus indicate what information channels would be required in real life. A further benefit from using a man-machine model is that one gains considerable insight into decision points and decision rules. Knowledge of these is important both for smooth operation of the simulation and as a guide to good decision rules in the real world.

However, we also felt that it was very important to be able to supplement the man-machine simulation with an all-machine model, since any simulation is essentially a statistical sample, and the number of runs of the model must be high enough to instill confidence in the statistical results. Because man-machine models are usually cumbersome, time-consuming and expensive to operate, an all-machine model is almost a prerequisite for a large number of runs.

Perhaps I should inject a word of caution here concerning the use of simulation to find optimal procedures. Since simulation is by nature a method of statistical sampling, there is usually no assurance, in a strict mathematical sense, that one has found an optimal procedure; however, one does have assurance that at least a feasible procedure has been found and that one knows its cost to a certain degree of accuracy. As often as not, in the real world, this is almost as valuable as knowing an optimal procedure; often, indeed, feasible procedures actually are close to optimal.

But to return to the description of our simulated missile squadron: since much of the missile hardware to be supported during a future time-period does not yet exist, it was necessary for us to make certain assumptions concerning equipment design, based on Atlas and Titan design data, discussions with various missile contractors, and RAND engineering estimates.

From these discussions, we decided that in order to simulate the hardware it would be necessary to identify the component parts only to the level required for studying supply, maintenance, and manning problems. It was not feasible for us to deal with the (approximately) 300,000 bits and pieces which comprise a ballistic missile system. Consequently, the level of identification adopted was defined as a "support unit". A "support unit" can be described as a replaceable part or "black box", or as a part that can be repaired in place, or a maintainable piece of ground-support or ground-handling equipment. Some judgment was required to determine which pieces of hardware were classifiable as "support units", but with proper engineering advice this was not too difficult. In this particular simulation, the final list of support units contained 1541 items, of which 786 comprised a missile, 261 were on ground support equipment, 33 on ground handling equipment, and 461 on other ground facilities.

To make the simulated hardware behave like a real system, it was necessary to collect much data about each support unit. This data can be broadly grouped under four headings: (1) reliability data, (2) supply data, (3) maintenance data, and (4) operational data. I should

mention that assembling data is perhaps the most important phase of any simulation design, particularly in simulating a system (such as a missile force) which is to come into being in the future. One reason is that we often have no other way to predict the behavior of a future system. We usually acquire data on system components long before we have data on the system as a whole. In the interim, simulation gives us a powerful tool for using known data in order to predict system characteristics.

In our case, we spent a great deal of effort and care in collecting what we hope and believe are correct data. For the reliability data, the developmental nature of the system made this task difficult, of course. We used two sources: (1) design requirements and engineering estimates, and (2) actual experience data from similar equipment (Nike, Bomarc, IOC Atlas, etc.). The numbers upon which we finally agreed, while not based upon actual usage experience (since none exists), represented a best set of estimates, and we have high confidence that they are at least coarsely accurate.

In order to use the reliability data in the planned operation of our simulation, we wrote the data into the following form. We hypothesized that every support unit would be subjected to one (and occasionally two) of the following stresses during every fifteen-minute period:

(1) Turn-on stress; resulting from power surge when a support unit is turned on;

(2) Operating stress; resulting from continuous power flow in a support unit -- usually less than the turn-on stress;

(3) Standing stress; the stress of time when a support unit is not being exercised in any way;

(4) Countdown stress; resulting from the pressurization and thermal shock of the fueling and loxing operations;

(5) Handling stress; due to the physical movement (transportation, etc.) of equipment; and,

(6) In-flight stress; caused by engine vibration and acceleration shock during launching.

Thus our reliability data was written in the form of a probability of malfunction per 15-minute increment, for each support unit and for each of the six stresses (in our case, 1541 support units times 6 stresses, for a total of 9246 failure probabilities).

To determine the supply data, we held what amounted to a parts provisioning conference. Using all the expert advice available to us, we went through the list of 1541 support units and determined or assigned the following items:

(1) Noun -- name of support unit

(2) Stock number -- for identification purposes

(3) Cost

(4) Weight

(5) Volume

(6) Expendability versus reparability

(7) Property Class

(8) Unit of issue

(9) Number of applications

(10) Criticality.

At the same conference, engineers and maintenance experts developed the data necessary to make our simulation behave realistically from a maintenance point of view. Consequently we developed the following maintenance data for each support unit:

(1) Lowest Level of Replacement: the lowest organizational level (launching location, blockhouse, squadron, depot, contractor) at which each support unit could be replaced, under various allotments of manning and resources.

(2) Lowest Level of Repair: the lowest organizational level at which each support unit could be repaired, under various allotments of manning and resources.

(3) Maintenance Requirements: the personnel, equipment resources and time necessary (by support unit) to remove and replace, repair in place, calibrate, diagnose malfunctions, and perform bench repairs.

(4) Time Between Overhaul and Time Between Replacement: Recognizing that periodic replacement or overhaul of certain items generates a significant work load, we determined which support units needed this periodic attention, the periodicity involved, and the resources required to accomplish the overhaul or replacement.

(5) Technical Order Compliance: Engineering modifications also generate a sizeable maintenance workload. We therefore decided that certain support units would be required to undergo a Technical Order Compliance during the simulation.

(6) Inspections: their kinds and frequencies, together with the lengths of time, personnel, and resources required.

The fourth set of data compiled for the simulation is that which we called the Operational or Status Data. An analysis was made of the various situations into which the major pieces of hardware of the missile system can be scheduled. Examples are: "Countdown of missile from T minus two-hours to T minus 15-minutes"; "Hold missile at 15-minute alert"; "Countdown from T minus 15-minutes and Launch"; "Daily, Weekly, Monthly Checkout"; "Installation of Missile in Silo"; "Engine Replacement"; etc. In all, approximately forty such situations were identified.

Each situation was divided into fifteen-minute time periods, and a determination was made as to what was done during each period. This, in turn, allowed us to determine which of the six stresses applied to each support unit during each fifteen-minute period of every situation. Going one step further and making use of the malfunction rates described earlier, we had the probability of malfunction of each support unit during each fifteen-minute time period of every situation. We were thus able to construct a mathematical failure model, which, by means of random draws, simulated the malfunctioning of missile system hardware undergoing almost any possible operational or maintenance situation.

I want to describe briefly the operation of the man-machine portion of the simulation. We decided to simulate a squadron which maintained a number of missiles on launchers in hardened silos and which held a number of missiles at squadron headquarters as spares. The launch complexes were theoretically separated a considerable distance from each other, and all complexes were theoretically located the same distance from squadron headquarters.

Six important management positions of the squadron were identified and staffed with Air Force Officers assigned from SAC. These positions were: (1) squadron commander and operations officer, (2) squadron maintenance officer, (3) squadron supply officer, and (4) three launch-complex officers. Each officer was furnished a clerical assistant to help with paper work. The squadron headquarters and each of the three launch complexes were located in different rooms; communications between these was only by telephone or by simulation mail service. Provision was made for monitoring the communications in order to be able to analyze the information flow.

IBM cards represented all of the remaining personnel of the simulated squadron and all of the resources. Thus, there was a card for every man in the Unit Manning Document, for every spare part in the Squadron Supply and Launch Complex stocks, and for every piece of test equipment and ground handling equipment. The management job of the participants was to allocate these resources to the various tasks that arose during the course of the simulation.

These tasks arose in the following fashion. A failure model, designed along the lines indicated earlier, made random draws to indicate whether or not a failure would occur for every support unit in the system, for each stress, and for every fifteen minutes of the proposed simulation. This failure model was run before the simulation began, and a scenario of its results were printed and given to the control clerks of the simulation. (The participants did not see this scenario.) As the simulation proceeded, the control clerks noted which support units would fail if subjected to a certain stress, and then checked to see if they were actually experiencing that stress at the time. If not, the stress was disregarded and there was no failure; if so, the clerks duly notified the participants of a malfunction in that particular support unit. It was then the obligation of the participant to collect the required resources in order to remedy the malfunction.

The participants and the control clerks communicated by telephone and by means of a console indicator box which signaled the status of each missile and whether or not a malfunction occurred. They worked in

different rooms to keep the participants from getting a sneak preview of the future.

I have briefly described the structure and contents of our man-machine simulation of a missile squadron. Actually, the model is much more sophisticated than it may appear to be in this short talk. For those interested, a more complete description is given in a RAND Classified paper S-107, "Laboratory Simulation of a Ballistic Missile Squadron: A Description of LP-II Phase 1.1", by William Haythorn.

Let me continue with a brief account of the events which trigger action in the system, as well as a short description of the simulation outputs.

The simulation was started with the presentation, to the squadron maintenance and supply officers, of lists of their available resources; to the three launch-complex officers, lists of their available resources and data on the current status of their missiles and other equipment; and to the squadron commander, a set of operational and maintenance requirements imposed upon the squadron by higher command. Thus, the squadron was told what operational program must be met, what performance standards must be achieved, what periodic maintenance schedules and what Technical Order Compliance requirements must be accomplished.

With the lists of available resources and operational requirements at hand, the operations, maintenance and supply officers jointly prepared a monthly operations schedule, and, from this, daily operations orders. The daily operations orders were sent to each launch complex, where the launch complex officers attempted to comply with them by initiating appropriate actions. These actions in turn caused malfunctions via the failure model, and the necessary malfunction correction procedures then began.

Finally, we imposed upon the simulation a set of external events which caused much activity. These were such events as severe weather conditions (e.g., a blizzard), certain catastrophic malfunctions, the imminence or actual occurrence of enemy action, or even the ultimate stress on the system imposed by a general war.

The outputs, or the information, which we received from the simulation consisted in general of answers to two of the questions which were implied in the beginning of this talk: (1) What is the operational effectiveness of the enterprise in meeting the objectives, with a given set of resource policies and a given operational program; and (2) What is the cost of the program?

We accomplished this by collecting and analyzing a great deal of data. For example, we recorded the status of each missile during every 15-minute period of the simulation, and the status of every resource (men and equipment) during every period. Thus, by analysis we were able to compare an achieved missile status posture with an attempted posture, which in itself is a measure of system effectiveness. Further, by analysis as well as by simulated firings, we were able to estimate the capacity of the squadron to put missiles on targets.

We calculated the costs by knowing, fairly accurately, the costs of the various resources which were put into the simulation. When these were added to the investment costs, we had a good estimate of the cost of the operation simulated, and were able to extrapolate from this to such other costs as were required (e.g., cost of a proposed missile force). Besides the cost, the resource-status data allowed us to determine resource utilization. We knew, for example, how much time a particular man or piece of equipment spent working on a job, in transit, standing by waiting for work, or holding up work by being occupied. These data were valuable for computing better manning and equipping tables.

I should like to add a few words concerning the all-machine model for making sufficient runs so that we have more statistical confidence in the results. At the time of writing this talk, we had not yet completed an all-machine model and so I am unable to give any description of what we have accomplished. I can, however, give a short outline of our plans and aims.

It became clear to us that even by using the largest computing machine available to the laboratory (the IBM 704) we could not begin to approach the complete mechanization of the simulation with the detail and richness that was used in the man-machine version. Consequently, we formulated a number of machine models or computations, each designed to simulate or shed light on one version of the overall simulation. Thus we designed a "prediction calculation" which showed the expected number of malfunctions resulting from any operational program. We built a "maintenance queuing model" to simulate the utilization of resources. We used a "stockage model" and a "cost calculation" which furnished the answers to those portions of the simulation.

We expect that our analysis of the results of these various models will afford statistical confidence in the overall answers. Also we are investigating the possibility of using the discovered relationships between the inputs and outputs of the various models and calculations in order to design a more inclusive missile squadron model which would react like the original man-machine model, although not with the same detail and completeness. However such a model will allow us to augment our investigation with the original model to furnish a very complete picture of the interrelationships among the variable parameters, and the cost and operational effectiveness of the system.

In summary, then, the conclusions of this discussion run as follows:

(1) The question of the value of reliability or of quality control is not a simple one, but rather must be considered in the light of any particular enterprise and its objectives.

(2) Simulation is frequently a powerful tool in answering this larger question of the complete relationship of reliability and quality control to the specific objectives of an enterprise.

(3) Simulation is frequently a useful method for using the known reliability of components of a system to determine the probable reliability of a proposed system as yet unbuilt.

(4) A useful simulation of any large complex enterprise must be very carefully designed, with particular attention that all essential factors and boundary conditions are present.

(5) One useful method for designing such a simulation is to use both a man-machine model and an all-machine model in concert -- the man-machine model to confirm the completeness of the simulation, and the all-machine model to supply the statistical confidence to be gained from extended runs.

LATENT RELIABILITY - AN APPROXIMATION METHOD

W. R. Davison
Phillips Petroleum Company

Introduction:

Reliability is the probability that any unit of product will work as the designer intended. Several mathematical models such as the Poisson and the Binomial are available for computing a probability-oriented estimate of reliability, but they require data which can be obtained only by extensive testing. Practical difficulties arise where performance tests destroy product of high unit value. Under certain conditions the Binomial calls for 3,000 or more tests; a less powerful approximation (such as one of the asymptotic functions discussed in the literature of mathematical statistics) might require as few as 300 tests. But when each test may cost $10,000 or more, even 300 is a prohibitive number.

This article introduces the latent reliability concept, an approach designed for product development situations where costly units are destroyed by testing. "Latent reliability" describes the agreement observed between engineering predictions of performance and actual performance, with respect to selected functioning characteristics. In essence, this concept equates the pooled percent accuracy of performance forecasts on units a, b, c, ... k, to the expected performance of the $(k + 1)^{th}$ unit. Since there is no inherent requirement for repeated trials of one design/process combination, the latent reliability approach lends itself particularly well to experimental work where each test may lead to design or process changes affecting the next test unit.

Discussion:

Military contracts covering development work on rocket motors generally contain a provision requiring the contractor to submit an estimate of end-item reliability prior to final qualification testing of the new rocket. The problem posed by this requirement differs from contract to contract depending upon factors such as the desired completion date, the size of the budget, and the scope of work to be done.

The typical rocket development project operates within fixed limits of time and money, often at a "crash program" tempo. If the unit undergoing development is a big one, the project engineer may be required to accomplish the transition from blueprint to blue yonder on the basis of 15 or 20 experimental firings. There is neither time nor cash to assemble and test several examples of each design. Not only do test models differ design-wise from each other but also the unit built for qualification (to which the reliability estimate must refer) may be several times larger than any model previously tested.

Reliability estimation boils down to "knowing," in the statistical sense, how a particular component is going to perform when subjected to test. This special knowledge is acquired by accumulating performance results until sufficient data are at hand to support a statement such

as: "There is less than 1 chance in 20 that the reliability of this product is below 0.98."

Probability-oriented estimates of reliability are available only through extensive testing performed either upon components, end-items, or both. Occasionally we find statements in the literature which leave the reader free to infer that reliability is a synonym for quality. Industrial experience runs counter to this belief. Quality, in the industrial sense, refers to compliance with applicable drawings and specifications, a condition notoriously independent of product performance. Countless engineering change orders have been and will continue to be initiated because parts made in conformance to the drawings failed to function satisfactorily in service. Furthermore, there is abundant evidence that an immature design or an excessively tight set of tolerances may produce a situation wherein non-conforming parts are found to work at least as well as those which meet specification requirements.

When product emanating from a given design/process combination is contaminated by the presence of a few unreliable units--say less than one percent--it is often difficult to discover that fact by sampling. The preponderance of good material makes it easy to draw a random sample comprised entirely of good pieces. Now if high reliability is urgently desired and if the performance test is non-destructive, bad pieces can be screened out of the lot by 100 percent testing. Where the test is destructive, however, economic restrictions limit the amount of testing and hence the amount of information available for reliability estimation. An impasse develops wherein the largest sample one can afford may be too small to detect the existence of material which cannot be tolerated. There is no escaping the fact that more information about reliability is gained from a single failure than from a hundred successes.

Where an optimum design and an optimum process can be assumed, end-item reliability may also be estimated from the individual component reliabilities. Suppose, for example, that the end-item is a simple assembly of three major components, and suppose it is known that the reliability of Component A is 0.99, the reliability of B is 0.94, and that of C is 0.96. Reliability of the end-item would then be $(0.99)(0.94)(0.96) = 0.89$, which illustrates the further point that nearly perfect components are needed to achieve a moderately reliable product.

It is reasonable to assert that the engineer designs purposely and selectively toward an optimum which he believes will meet a given set of performance requirements. Changes may be introduced to compensate for faults revealed by static tests, but each modification is a deliberate move in the direction of a fully adequate design. As development work proceeds, the designer can forecast, within fairly close limits, what effect each modification will produce upon the various performance characteristics. If, therefore, in k trials of a given rocket type, the observed values of, say, Effective Impulse fall within some predetermined allowable range of forecast values, it can be argued that the $(k + 1)^{th}$ forecast of Effective Impulse is also likely to be verified by the observed test value.

The designer's ability to satisfy a given set of performance requirements can be measured by comparing forecast versus actual performance in successive trials of related units, and the results then imputed to the expected performance of the next unit tested. Such measurements obtained for each performance characteristic can be combined to yield a score which describes the over-all degree of agreement between predictions and observed performance. This over-all score is termed the "latent reliability" of the design.

Restrictions:

The rationale of latent reliability applies to a series of units produced within the same program and should not be used to embrace a number of different projects each having its own set of design requirements or objectives. Thus it would be improper to attribute to the first unit fabricated under Project D the latent reliability achieved in connection with Project A, B, or C.

An additional restriction affects the kind of evidence which may be used for scoring purposes. To a degree, we beg the fundamental question of reliability by relating the estimate to a comparison of actual versus predicted performance, for even if the predictions are verified by test results, the observed performance may still be inadequate. In other words, it would be possible to produce an estimate of perfect latent reliability on the basis of accurately predicted failure. The procedure therefore requires success forecasts and assumes that the model specification (to which qualification rounds are designed and fabricated) will specify performance characteristics acceptable to the sponsoring agency or service.

Finally, it is desirable to correct for the number of tests in order that scores will tend to be conservative. The recommended correction factor, $-100\ (0.5^k)$, is applied to the pooled score as explained below. Thus the highest value possible for latent reliability estimates based upon a single test is 50; i.e., 100 - 100 (0.5). If the first test failed completely, use of the correction factor would produce a negative value of the latent reliability estimate. Negative values are therefore defined as zero.

Computation:

The procedure for developing a latent reliability estimate is presented with reference to Table I, using data on the performance of a fictitious rocket. Modifications of the method shown might be required for other applications.

Step 1. List the performance characteristics of interest. These characteristics are dictated by the design objectives of the unit undergoing development.

Step 2. Weight each characteristic according to its importance in the particular experimental unit to be tested. The sum of weights equals 100.

Step 3. Prior to test obtain the forecast value of each characteristic from the project engineer.

Step 4. Complete the record by entering test results.

Step 5. Compute p_a, the measure of relative agreement between forecast and observed values for each characteristic.

$$p_a = 1 - \frac{F_p - F_o}{F_p},$$

where F_p and F_o are, respectively, predicted and observed values. When $F_p = F_o$, p_a takes its maximum value, i.e., $p_a = 1$. Negative values of p_a, which arise when $F_o > 2 F_p$, are treated as zero. Consideration was given to the use of penalties or negative scores in the case $F_o > 2 F_p$. This notion was abandoned because characteristics listed in Table I are not mutually independent. Large differences between forecast and observed results for one characteristic are, in the present application, reflected in comparable differences elsewhere. In other words, the penalty system is built-in, although this might not be the case in some other development situation.

Step 6. Multiply each p_a by its corresponding weight, W.

Step 7. Compute the raw score, $\Sigma\, p_{a_i} W_i = (p_{a_1} W_1) + (p_{a_2} W_2) + \ldots (p_{a_n} W_n)$.

Step 8. The raw score from Step 7 can now be combined with those of earlier tests in the same series by computing the over-all average.

Step 9. Apply the correction term, $-100\ (0.5^k)$, to the pooled raw score computed in Step 8. The exponent k represents the total number of tests in the series. If $k > 7$ the correction factor becomes negligible and may be omitted. The result of this step estimates latent reliability applicable to the forecast parameters of the next unit in the series to be tested.

Similarly, the latent reliability estimate for a single characteristic can be computed over units in the same or closely related series of tests. For example, if interest were centered upon Duration (See Table I), the unweighted p_a values for that variable may be pooled either by computing an arithmetic mean or by taking an over-all product and extracting the k^{th} root to obtain the geometric mean, p_a^*. (If any $p_a = 0$, the arithmetic rather than the geometric mean is used.) The correction terms -0.5^k should then be applied, since the average used is a decimal fraction.

In defining a test series, recognition is given to the progressive complexity and importance assumed by experimental test units as a development project moves toward completion. Early test sequences often involve scaled-down models, or may be devoted primarily to an investigation of one design characteristic, and are thus less important than a later sequence or test series employing large scale units. These differences, where they arise, are accounted for by assigning weights to each test series before computing the grand average latent reliability which will be attributed to the qualification test rounds.

Under the suggested procedure, the corrected latent reliability value of each series is multiplied by the corresponding series weight factor, after which the mean is obtained as outlined in Step 7. (Where the geometric mean is used, division of the over-all product by the sum of weights is performed before the k^{th} root is extracted.)

Conclusion and Summary:

While the validity of using forecast versus observed performance to estimate reliability is not subject to rigorous proof, the method outlined above yields a meaningful and usable estimate of progress in a development program toward an adequate design/process combination. The latent reliability concept is applicable to situations wherein extensive performance testing is not feasible, but should not be considered an equivalent substitute for reliability estimates from the Binomial or other probability-oriented models.

Bibliography:

1. "Evaluation of Reliability in Product Development" by Myron Lipow and Richard D. Geckler, American Rocket Society reprint; paper read at ARS Fall Meeting, 1955, Los Angeles, California.

2. Lusser, Robert, Reliability of Guided Missiles, Redstone Arsenal, Huntsville, Ala., 1954.

3. Lusser, Robert, Proposed Reliability Specifications for Guided Missiles, Redstone Arsenal, Huntsville, Ala., 1955.

4. Shewhart, W. A., Economic Control of Manufactured Product, D. Van Nostrand, Inc., New York, 1931.

5. Gore, W. L., Statistical Methods for Chemical Experimentation, Interscience Publishers, Inc., New York, 1952.

Acknowledgement:

The author acknowledges his debt to Dr. Edwin F. Moore* for valuable contributions to the scoring method and procedure discussed in this article.

* Now with ASTRODYNE, Inc.

TABLE I--COMPUTATION OF LATENT RELIABILITY

Firing: Round 3

1	2	3	4	5	6	7	8
Characteristic	Weight (W)	Predicted (F_p)	Observed (F_o)	Col 3 Minus (a) Col 4	Col 5 Div. by Col 3	1 minus Col 6 (p_a)	Col 7 times Col 2
Ignition Delay (ms)	5	300	500	200	.67	.33	1.65
Duration (Tangent time)	50	90	75	15	.17	.83	41.50
Initial Pressure	3	250	240	10	.04	.96	2.88
Final Pressure	3	235	300	65	.28	.72	2.16
Maximum Pressure	5	260	520	260	1.00	-0-	-0-
Minimum Pressure	3	230	240	10	.04	.96	2.88
Average Pressure	6	250	420	170	.68	.32	1.92
Initial Thrust	3	2,980	3,420	440	.16	.84	2.52
Final Thrust	1	2,840	3,600	760	.27	.73	.73
Maximum Thrust	2	3,110	4,750	1,640	.53	.47	.94
Minimum Thrust	1	2,760	3,420	660	.24	.76	.76
Average Thrust	3	2,955	3,700	745	.25	.75	2.25
Total Impulse	4	266,000	285,000	19,000	.07	.93	3.72
Tangent Impulse	6	248,700	221,200	27,500	.11	.89	5.35
Specific Impulse	2	140	143	3	.02	.98	1.96
c*	3	3,300	3,260	40	.01	.99	2.97
Totals	100	...	...	...	...	...	74.19

Raw Scores: This round 74.19
Round 2 ..... 64.29
Round 1 ..... 66.71

(a)Absolute value

Average Raw Score = 74.19 + 64.29 + 66.71 ÷ 3 = 68.39
Correction Factor: -100 (.53) = -12.5
Latent Reliability: 68.39 - 12.5 = 55.89 or 56%

MACHINE CAPABILITY - A TOOL FOR BUILDER AND USER

David N. Smith, Manager, Research & Development
Jones & Lamson Machine Company

It is undeniably so that the use of analytical techniques to establish the exact nature of a machine's performance has been receiving increasing attention in recent years, from both the machine builder and the machine user. It is also certain that machine capability is so powerful and exact a tool that interest in, and use of it, are bound to grow in years to come.

The measurement of machine capability applies the science of statistics to process variations. No production process ever produces pieces that are exactly alike. There is some inherent variability which can never be eliminated. These inherent variations are usually small in magnitude but are large in number. The total amount of basic variability in a given machine tool process depends upon such uncontrollable variables as piece to piece hardness, environmental temperatures, minute deflections in the machine, periodic machine vibrations, runout in the spindle bearings, etc. All of these variables combine in a random manner to cause small variations in the product. Such variations, therefore, result from a large number of contributory causes, no one of which is comparable to the combined effect of all and which are just as likely to produce positive as negative deviations from the mean of all the measurements taken. A study of the statistical properties of these variations shows that they follow certain basic laws described by the theory of probability.

A machine tool which is functioning within the mechanical capabilities built into it and which is not being influenced by assignable causes of variation will produce parts which show these small measurement variations. If a plot is made with measurement deviation from the mean as the abscissa and the frequency of occurrence of these deviations as the ordinate, the resulting histogram will approximate the so-called Normal Distribution Law. A curve fitted to this histogram will be characteristic for a particular process on a specific machine. It is the dimension of this normal curve which defines the machine capability.

More specifically, machine capability is defined as the minimum spread of a specific measurement variation which will include 99.7% of all the measurements from the given process. This 99.7% figure arises from the statistical fact that a spread of ± 3 standard deviations from the mean will include 99.7% of all the measurements, if these measurements are normally distributed about their mean. Therefore, an average of only 3 pieces in 1000 can be expected to have measurements which fall outside these minimum limits. This standard total process width of 6 standard deviations is used universally throughout American industry.

In measuring machine capability for a particular process, it is necessary to study the statistical properties of the measurement data and to determine such basic statistics as arithmetic mean, variance, standard deviation, etc. and to make the proper corrections in the data so that the final answer will not include any measurement variation for which there is an assignable cause. Various statistical tests are

useful in helping to make decisions regarding the presence or absence of assignable causes of measurement variation. These tests must be applied correctly and to the proper statistical situations or incorrect conclusions will be drawn from the results.

In conducting a machine capability test, it is essential to control carefully the actual machining test so that all known causes of measurement variation are eliminated. This means that the pieces to be machined must be uniform with respect to size, material, finish, etc. The machine itself must operate at as uniform a temperature as possible, machine vibrations must be minimized, and no manual adjustments made to the process or machine while the test is being run. In other words, the machine environment must be kept as nearly constant as possible and care must be taken to insure that no known outside disturbances are allowed to affect the machining process.

All measurements should be taken with a gage which has at least one order of resolution higher than that required by the usual run of work from the machine.

The objective of a machine capability test should be to arrive at the minimum process width which is defensible for the machine in its present state of design.

Before any computations relating directly to the machine capability are made, it is necessary to test the process to see that it is in statistical control. Lack of statistical control throughout the data indicates that assignable causes of measurement variation still exist and that the conditions under which the measurements were taken may be unsatisfactory and more careful control of the test is required. In many situations it is necessary to exercise considerable judgment concerning the tests for statistical control because borderline cases may occur.

Machine tool capability tests usually show trends in the data due to a known but non-correctable cause such as tool or diamond wear. Statistical techniques exist for removing such trends from the data.

In a carefully controlled machine capability test where the number of pieces is sufficiently large (not less than 50), it is proper to assume that the measurements taken are in themselves a universe of parts. This assumption then puts any discussion of the test results in terms of the actual pieces measured. Statistical estimates are then avoided and the final machine capability measurement is easier to explain and defend.

None of the required calculations for determining machine capability involve anything more than ordinary arithmetic. However, the amount of calculations can become large and therefore it is important to use all the legitimate short cuts possible in order to reduce the amount of labor.

There are two general situations commonly met in machine capability measurements. The first is the case in which the measurement data show no discernible trend. The second case is that in where there is a significant trend, as from tool wear, and this should be dealt with in a manner different from the first case, by correcting the data to remove

the trend, "levelling" the data. This is simply (though tediously) done by fitting a computed trend line to the raw data by the least-squares method, and correcting each point in the data sequence by the amount of the rise (or fall) of the trend line at that point. (Note the word "computed". A trend line established by a visual fit is generally not good enough). This eliminates the effect of a known assignable cause of variation in the process, which we cannot eliminate in actual practice, but which is not a true part of the basic machine performance.

Before a value for machine capability is finally stated, the data must be given several checks for validity. It must be tested for randomness, which includes testing for number of runs above and below the median, the number of runs up and down, and the mean square successive difference test. These tests are all described in published literature on this subject. These tests for randomness will establish that the machine test was or was not free from disturbances resulting in either long term trends, or short rapid oscillations, in which case we must look for and try to eliminate the causes for such disturbances.

The data should also be checked by making a control chart for averages and a control chart for range. If the data meets all these tests, then we are confident that the data came from a process that was in statistical control, and we can compute and state the value of machine capability with corresponding confidence and exactness.

Let us now look at the usefulness of this technique from the standpoint, first of the machine tool builder, and next of the machine tool user; to see how it can help us to achieve what each desire: improved end product coming off the machine.

The machine tool builder's first concern is in making sure that his product can meet required specifications, and that it will satisfactorily meet competition. But what do we mean by "specifications"? Not, in this case, swing, speed, nor horsepower, which are easily and directly measurable, but, rather, the dimensional quality of work pieces produced. This is just what machine capability tells us, with precision and certainty, making possible the offering of a machine to a customer with assurance that it will meet his requirements.

The builder is also concerned with improving his product. He hopes a new model will be better than the current model, and how else can he determine this, as far as quality of output is concerned, than by this quantitative yardstick of machine capability?

Thirdly, and increasingly so as our machines get more and more automated, the builder may wish to make the machine self-regulating; automatically to measure its dimensional output, and apply feed-back signals to the tool setting to compensate for any trends in the process. To do this successfully, and to design a proper regulating system, he must have exact knowledge of the machine capability, the sources of possible disturbances, and the rate of expected trends. This imposes on the control system designer an entirely new problem, and lands him squarely in the middle of the machine capability concept.

Now, from the buyer-and user's point of view, what does a knowledge of machine capability do for him? His first concern is for buying

equipment which will meet his product-quality requirements, without producing an objectionable percentage of scrap, at a desired production rate, and without the need for exceptional operator skills to compensate for machine deficiencies. (We all remember stories of the old-time machinist who could make an old worn-out machine jump through hoops and turn out precision work, but they are hard to come by these days). He wants a comfortable margin between his work-piece tolerances and the expected performance of the machine.

But the buyer is also just as much concerned in avoiding "over-specifying" the machine, and buying one with a much greater capability than he needs, at a correspondingly high cost. Does he need a drill press or a jig-borer? Here, a knowledge of machine capability gives both buyer and builder a common language, and a specification tool which will assure each that the best buy is being made.

From the user's standpoint, a knowledge of machine capability provides him with another interesting tool. During its useful life, all machines are subject to wear and degradation. If the quality of product produced deteriorates to an objectionably low level, a new test for machine capability will tell the user whether the fault lies in the machine, or in some change in its mode of operation. In the first case a reconditioning or replacement of the machine would be indicated. In the second case, the user would look for and correct the cause for the deficiency within his own operation, and avoid the expense and the interruption of a repair or rebuilding job.

We are currently witnessing the advent of a new type of equipment - the numerically controlled machine tool. With these machines the information to be fed into the machine, generally from punched or magnetic tape, is prepared in the office, the tools are accurately preset before being placed in the machine, and the hope is that the first piece made will be a good piece. But without thorough knowledge of the capability of what is, in this case, a system rather than a conventional machine, what chance would there be of achieving this? In fact, the designer of the control system has to have complete knowledge of the machine capabilities in many characteristics before he can design the controls to meet specifications, and have the machine and the controls compatible.

It is hardly necessary to point out that a knowledge of machine capability is indispensible to the quality control engineer. It is his principal stock-in-trade. This knowledge is also of great usefulness to both parties in any sub-contracting operation, as a protection to the placer of the sub-contract that he will get what he wants at a fair price, and to the bidder on the sub-contract that he can do what is asked of him at a reasonable profit.

Machine capability can thus be looked upon as an extremely useful "common-language" tool; useful to the builder, the buyer, the machine sales engineer, the user, the quality-control engineer, and even to the operator, whether or not he is aware of it. We can expect to see it appearing with increasing frequency in purchase specifications, and we can expect to see it appearing with increasing frequency in machine advertising. It is an exact yardstick, and thus important and valuable.

QUALITY CONTROL IN RUBBER MANUFACTURING

Dr. J. N. Berrettoni
Chairman, Statistics Department
Western Reserve University

This paper presents illustrations of quality control techniques that actually were used to formulate decisions in a rubber job shop. The format of the paper is based on the experiences of the Quality Control Engineer (QCE) in solving by statistical quality control some of the daily problems of the chief inspector, rubber technologist, engineer, salesman, and plant manager. Instead of presenting the common illustrations of $\bar{X}$ and R charts, p and c charts, and attribute sampling plans this paper concentrates on:

1. Variable sampling plans based on the range for disposition of a lot given a specified AQL - tensile strength.

2. The t-test for disposition of a lot given a lot average - tensile strength.

3. Confidence interval for a lot average - tensile strength.

4. The F and t-tests for selecting the better rubber compound between A and B - flex life.

5. Determining one sided statistical tolerance limits - flex life.

6. Statistical addition of weight variation for different lengths of unvulcanized rubber blanks.

7. Statistical control of group weighing - spherical rubber centers of golf balls.

8. Statistical elimination of machine differences by correlation so that lot disposition is based on product quality variation rather than machine variation - flex life.

9. "Stiffness evaluation" of single and double variable sampling plans by means of the non-central t-distribution - flex life and the 20% Dry Compression test.

10. Analysis of variance of linear, quadratic, cubic and quartic trend effects of a mold for eccentricity values of rubber parts.

11. The statistical theory and control of batch blending - a statistical measure of blend uniformity at each blend.

12. Reciprocal transformation of flex life in order to obtain a normal distribution.

The inspector was looking over a lot of several thousand rubber bushings which had been screened for cracks, bursts and non-fills. The only decision remaining was that of accepting or rejecting the lot for tensile quality. The inspector remembered two things: 1) the blueprint called out a minimum specification of 1825 psi and 2) the buyer had told him specifically one day that no lot in excess of one percent below 1825 psi would be acceptable. In other words, the acceptable quality level (AQL) of tensile was 1%.

The inspector decided to base his decision whether he should accept or reject the lot on the results of a single sample. Habit and precedence dictated a sample size of ten. Having priority on testing equipment, the inspector immediately obtained the following tensile values in terms of psi. He noticed quickly that none of the values was below 1825 psi and therefore, the sample percent defective was zero. He OK'd the lot

1920, 1840, 1970, 1940, 1860, 2060, 2030, 1840, 1990, 2170

for shipment. Having made this decision he began to wonder about its reliability. He even wondered about the chances of obtaining a zero percent sample from a lot in excess of 1% non-conformance. Had he made a good or a bad decision - would the lot be returned?

Somewhat bothered about the uncertainty of his decision, the inspector decided to hold up shipment and to find out from the company's Quality Control Engineer whether the lot met the minimum specification of 1825 psi and an AQL of 1%. The first thing the QCE did was to stress the certainty of his forthcoming decision. He said that he would analyze the sample data so that if the lot is 1% or less in non-conformance the certainty of accepting the lot would be 95 in 100. He used the following variable statistical sampling plan (1) to obtain this degree of certainty.

1. Group the sample results into subgroups of five. For the above sample size of ten we have:

GROUP I	GROUP II
1920	2060
1840	2030
1970	1840
1940	1990
1860	2170

2. Determine the range value for each group.

GROUP I: Range = 1970 - 1840 = 130
GROUP II: Range = 2170 - 1840 = 330

3. Calculate the average range and symbolize it by $\bar{R}$.

$$\bar{R} = \frac{130 + 330}{2} = 230$$

4. Find the average tensile of all ten values and symbolize it by $\bar{X}$.

$$\bar{X} = \frac{19620}{10} = 1962$$

5. Choose from the following Table I the "k" factor corresponding to an AQL of 1%. This table gives an assurance of at least 95 in 100 of accepting a lot with an AQL of 1% or less.

TABLE I

k FACTORS CORRESPONDING TO GIVEN AQL'S
FOR SAMPLE SIZE TEN

k	AQL (%)
1.0	4.6
1.1	3.5
1.2	2.6
1.3	1.8
1.4	1.3
1.5	1.0
1.6	0.7
1.7	0.4

k equals 1.5 for an AQL of 1.0%

6. Calculate the quantity

$$Q = \bar{X} - \frac{(k)\ (\bar{R})}{2.3259} = 1962 - \frac{(1.5)(230)}{2.3259} = 1812 \text{ psi}$$

where 2.3259 is a constant reflecting the subgroup size of 5.

7. Compare the Q-value to the lower specification (L). When

   a) Q is equal to or greater than L conclude that the lot meets the AQL and therefore accept the lot.

   b) Q is less than L conclude that the lot is in excess of the AQL and therefore reject the lot.

8. Conclusion: Since 1812 psi is less than 1825 psi conclude that lot is in excess of 1% non-conformance to 1825 psi minimum and is therefore rejected.

The QCE had no misgivings about his decision because he knew the statistical plan accepts a lot by at least a 95 to 100 chance if the lot is 1% or less in non-conformance. And this assurance still holds in spite of the fact that the sample shows zero percent below 1825 psi.

The inspector was now thoroughly convinced that the lot of bushings did not meet an AQL of 1%. However, a great deal of money had been poured into making the rubber bushings and he wondered what could be done with the lot before deciding to relegate it to the scrap barrels. Screening for tensile was out of the question. Again, he remembered two things: 1) another customer would be willing to accept the lot if the lot average was 2000 psi and 2) the rubber technologist had told him that the compound for this part would yield a lot average of 2000 psi. The inspector felt the lot could be accepted for shipment to this customer. However, there was one disturbing fact - the sample average was 1962 psi - 38 psi below 2000 psi. He thought and wondered whether the lot average was actually 2000 psi and the low sample average of 1962 psi was due to sampling variation - the inevitable variation resulting in taking a sample. If this were so the lot should be accepted and shipped. Concurrently he also wondered whether the smaller sample average reflected not a sample variation but a smaller lot average than 2000 psi. Something may have happened to the compound or some change may have taken place in processing to yield a lot average smaller than 2000 psi. If this were the case the lot should be rejected. The inspector was at a loss for an answer and once again called upon the QCE for assistance.

The QCE used the following well known statistical method to obtain an answer.

1. Determine the average of the sample

$$\bar{X} = \frac{19620}{10} = 1962 \text{ psi}$$

2. Determine the standard deviation of the sample. This is a measure of variation which uses the arithmetic mean as a point of reference. Its formula is

$$S = \sqrt{\frac{\sum (X - \bar{X})^2}{n - 1}}$$

where X and n refer to the individual sample values and number of values (or sample size) respectively. For the ten tensile values

$$S = 105.8 \text{ psi}$$

3. Calculate the quantity

$$t = \frac{(2000 - \bar{X})\sqrt{n}}{S} = \frac{(2000 - 1962)\sqrt{10}}{105.8} = 1.13$$

4. Choose from the following Table II a $t_{.05}$ value corresponding to the number n - 1. This number of n - 1 is called degrees of freedom and "df" is used for its abbreviation.

5. Compare the calculated t and the tabulated $t_{.05}$. Then, when

a) .t equals or is less than $t_{.05}$ conclude that the lot has an average of 2000 psi and therefore accept the lot

b) t is greater than $t_{.05}$ conclude that the lot has an average less than 2000 psi and therefore reject the lot.

TABLE II

Degrees of Freedom	$t_{.05}$
1	6.31
2	2.92
3	2.35
4	2.13
5	2.02
6	1.94
7	1.90
8	1.86
9	1.83
10	1.81

6. Conclusion: Table II shows that $t_{.05}$ for 9 degrees of freedom is 1.83. Since t of 1.13 is less than a $t_{.05}$ of 1.83, conclude that the lot average is 2000 psi and the lot is therefore accepted.

The QCE felt rather confident about this decision because he told the inspector the tabular $t_{.05}$ values gave an assurance of 95 in 100 of being correct when such a conclusion is made. Needless to say, the inspector immediately shipped the lot.

Later on that day and after the lot was well on its way, the inspector telephoned the QCE. He said there was another disturbing fact about the 2000 psi that he had forgotten to mention at the time. It was this: there was no past data to substantiate the technologist's assertion that the compound would yield an average of 2000 psi. The inspector wondered whether the decision to accept the lot still held. The QCE said he didn't know. In reality he would have to use a different method of analysis in order to arrive at a more realistic decision. The method now is:

1. Calculate the mean of the sample, $\bar{X}$ = 1962 psi

2. Calculate the standard deviation of the sample, S = 105.8 psi

3. Choose $t_{.05}$ = 1.83

4. Determine the two values

$$\bar{X} + \frac{(t_{.05})(S)}{\sqrt{n}} = 1962 + \frac{(1.83)(105.8)}{\sqrt{10}} = 2023 \text{ psi}$$

$$\bar{X} - \frac{(t_{.05})(s)}{\sqrt{n}} = 1962 - \frac{(1.83)(105.8)}{\sqrt{10}} = 1901 \text{ psi}$$

5. Conclusion: There is a 90% confidence in the conclusion that the lot average is some one value in the interval 1901 to 2023 psi.

The QCE noticed the interval of 1901 to 2023 psi included a range of values less than 2000 psi. This meant the lot average could be less than 2000 psi and therefore acceptance of lot would be denied. The decision to accept the lot no longer held. The only recourse to the inspector and QCE was to hope that 1) the technologist knew what he was talking about when he made the statement the compound would yield an average of 2000 psi or 2) the customer would overlook this specification requirement or better yet change it to 1900 psi.

The QCE decided to drop by the laboratory and ask the technologist about the average yield of the bushing compound. The technologist reaffirmed his belief that the compound which he personally had formulated would yield an average of 2000 psi. The technologist then mentioned he had heard of the good work of the QCE and coincidentally wanted some help himself. The other day he had formulated two new compounds, A and B. He wanted to know which compound gave the better flex life where

a) less variation, and
b) higher average

spelled out the better compound. He already had a good idea that B was the better compound because the sample results of Table III shows B with less variation (45.66 for B to 71.0 for A) and a higher average (293 for B to 270 for A). In addition this superiority was expected because com-

TABLE III

FLEX LIFE IN KILOCYCLES, FOUR SPECIMENS

	SAMPLE A	SAMPLE B
	179	323
	343	230
	253	289
	305	330
Average $\bar{X}$	270	293
Variation S	71.0	45.66

pound B had more expensive ingredients and involved a longer curing time. However the technologist didn't want to accept the sample results at their face values because in the past he had made some serious errors in doing so. He wanted to know what the QCE could do to help him in making the

decision whether B was better than A.

The QCE proceeded to outline the steps of the statistical method appropriate for this problem.

I. Determine Whether Variation of Compound B is Actually Less Than That of A.

1. Obtain the square of the standard deviations. The square of a standard deviation is called variance.

   $S_A^2 = 5041$ with $(n_A - 1)$ or 3 degrees of freedom

   $S_B^2 = 2085$ with $(n_B - 1)$ or 3 degrees of freedom

2. Divide the larger variance by the smaller and designate the variance ratio by the letter F.

$$F = \frac{5041}{2085} = 2.42$$

3. Compare the calculated F with the $F_{.05}$ obtained from Table IV for $(n_A - 1)$ and $(n_B - 1)$ degrees of freedom.

TABLE IV

VARIANCE RATIOS - $F_{.05}$

Degrees of Freedom Smaller Variance	Degrees of Freedom - Larger Variance				
	1	2	3	4	5
1	164.4	199.5	215.7	224.6	230.2
2	18.5	19.2	19.2	19.3	19.3
3	10.1	9.6	9.3	9.1	9.0
4	7.7	6.9	6.6	6.4	6.3
5	6.6	5.8	5.4	5.2	5.1

When the F value is equal to or less than $F_{.05}$, then conclude the variation of both compounds is the same even though the sample results show otherwise. When the F value is greater than $F_{.05}$, then conclude the variation of compound A is greater than that of B.

4. Conclusion: Since $F = 2.42$ is less than $F_{.05} = 9.3$ conclude there is no difference in compound variation; that is the difference between 45.66 and 71.0 is due to

sampling variation and not to a difference in compound ingredients or longer curing time. As there is no difference in variation, the two values are combined to obtain one value of variation. This combination must be done by keeping in tact the mean reference point of each sample. Formula wise, we have

$$S = \sqrt{\frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{(n_A - 1) + (n_B - 1)}} = \sqrt{3563} = 59.69$$

II. Determine Whether the Average of Compound B is Actually Greater Than That of A.

1. Determine the difference between the averages

$$\bar{X}_B - \bar{X}_A = 293 - 270 = 23$$

2. Calculate

$$t = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{S^2}{n_A} + \frac{S^2}{n_B}}} = \frac{23}{\sqrt{1781.5}} = \frac{23.0}{42.2} = .54$$

3. Compare the calculated value of t with $t_{.05}$ for $n_A + n_B - 2$ degrees of freedom (Table II), when t is equal to or less than $t_{.05}$ conclude there is no difference in compound averages. When t is greater than $t_{.05}$ conclude there is a difference.

4. Conclusion: Since t = .54 and $t_{.05}$ = 1.94 the conclusion is that there is no difference in average yield between compounds A and B. As a result one average is determined.

$$\bar{X} = \frac{\bar{X}_A + \bar{X}_B}{2} = \frac{270 + 293}{2} = 281.5$$

Feeling somewhat exhausted by the above ordeal of calculations, the QCE turned to the technologist and informed him that there was no difference in the averages or the variation of the two compounds. In addition he reiterated the assurance of 95 in 100 of accepting a no difference decision when it was actually true that there was no difference between the two compounds.

The technologist was rather astounded at the no difference conclusions, especially the variation difference. However, the 95 in 100 assurance pleased him somewhat and he decided to go along with the QCE. However, he still felt himself at somewhat of a loss because he didn't know how to solve his assigned task of establishing a lower tolerance or speci-

fication limit for this new part. And, as it is to be inevitable in these discussions, he turned to the QCE for guidance and assistance. The QCE once more explained what to do in order to obtain a reliable determination (2).

1. Calculate the value

$$\bar{X} - KS = 281.5 - (2.755)(59.69) = 117$$

where K is given in Table V for given sample size, n.

2. Conclusion: Assuming a normal distribution, the factor K gives the chance of 90 in 100 that at most 5% of production will be less than a lower tolerance limit of 117.

TABLE V

K - FACTORS FOR DETERMINATION OF
ONE - SIDED STATISTICAL TOLERANCE LIMITS
$\bar{X} - KS$ OR $\bar{X} + KS$

n	K
3	5.310
4	3.957
5	3.400
6	3.091
7	2.894
8	2.755
9	2.649
10	2.568
11	2.503

The technologist was not too pleased with such a low value of 117 kilocycles and immediately began to alter the proportions of the ingredients - more carbon black - a smaller amount of plasticizer. While this thinking was taking place, the QCE left for his office.

Upon entering his office the QCE was greeted by a young engineer who had several questions to ask of him. He had just taken a sample of fifty 36-inch lengths of unvulcanized rubber blanks from the continuous extruded rubber compound of a tube machine. He used a yardstick and a pair of scissors to obtain his sample of fifty 36-inch lengths. The sample of fifty 36-inch blanks yielded an average weight of 826 grams ($\bar{X}_{36}$) and a variation, standard deviation, of 17.5 grams (S_{36}). Now, the young engineer wanted to know what the average and standard deviation would be for 1½-inch blanks because this was the size to be used in their multicavity molds. Would he have to cut each 36-inch length into twenty-four 1½-inch lengths? This would give a total of twelve hundred 1½-inch blanks. Also it would entail eleven hundred and fifty cuttings. This, to say the least, would be laborious. So he wondered whether he could merely divide the average and standard deviation of the 36-inch blanks by the number of times 1½ goes into 36 (i.e. 24) in order to obtain the

values of the average and standard deviation for $1\frac{1}{2}$-inch blanks. Obviously this arithmetical manipulation would save a lot of time and effort. The QCE answered as follows. It wasn't necessary to cut to $1\frac{1}{2}$-inch lengths. The average weight of $1\frac{1}{2}$-inch blanks could be obtained by dividing the average weight of the 36-inch blanks by 24. But, in order to obtain the standard deviation of $1\frac{1}{2}$-inch blanks, the standard deviation of the 36-inch blanks must be divided by the square-root of 24. That is,

$$\bar{X}_{1\frac{1}{2}} = \frac{\bar{X}_{36}}{24} = \frac{826}{24} = 34.42 \text{ grams}$$

$$S_{1\frac{1}{2}} = \frac{S_{36}}{\sqrt{24}} = \frac{17.5}{4.9} = 3.57 \text{ grams}$$

By assuming normality so that $\bar{X} - 3S$ can be used to set tolerance limits, the QCE pointed out that when 36-inch blanks vary from

$$\bar{X}_{36} - 3S_{36} = 773.5 \text{ (low)} \text{ to } \bar{X}_{36} + 3S_{36} = 878.5 \text{ (high)}$$

then, the $1\frac{1}{2}$-inch blanks will vary from

$$\bar{X}_{1\frac{1}{2}} - 3S_{1\frac{1}{2}} = 23.71 \text{ (low)} \text{ to } \bar{X}_{1\frac{1}{2}} + 3S_{1\frac{1}{2}} = 45.13 \text{ (high)}$$

Or, if specification limits for $1\frac{1}{2}$-inch blanks are 22 (L) and 40 (U) grams, then

$$d_{X_U} = \frac{U - \bar{X}_{1\frac{1}{2}}}{S_{1\frac{1}{2}}} = \frac{40 - 34.43}{3.57} = \frac{+ 5.58}{3.57} = + 1.56$$

$$\therefore X_U = 5.94\% \text{ (Normal Table)}$$

$$d_{X_L} = \frac{L - \bar{X}_{1\frac{1}{2}}}{S_{1\frac{1}{2}}} = \frac{22 - 34.42}{3.57} = \frac{- 12.42}{3.57} = - 3.48$$

$$\therefore X_L = .025\% \text{ (Normal Table)}$$

and, therefore, 5.94% of the $1\frac{1}{2}$-inch blanks will be in excess of the upper specification of 40 grams. The lower specification is pretty well satisfied. The obvious correction in this instance would be to lower the process average from 34.42 grams to the specified nominal of 31 grams. Then only six tenths of one percent would be outside either limit. The young engineer was grateful for this information and especially appreciated the intricate relationship existing between averages and variation for different length blanks. Upon leaving the QCE's office he headed directly for the shop because he was rather curious to find out whether the oversize blanks were responsible for a rash of excess flash which had

occurred lately.

This problem on weight of blanks reminded the QCE that similar reasoning could be applied to the control of weight of spherical rubber centers of golf balls. The control in this case would be based on weighing in groups because usually where a weight specification is to be met on small parts it will be advantageous to weigh samples for quality control a few at a time rather than to weigh each individual part. Weighing in groups, say, five at a time, requires 1/5 the number of weighings needed for individuals and yields more weighing accuracy.* Or with the same number of weighings used for individuals the five-fold increase in sample size provides a considerable increase in sampling accuracy.

There was considerable statistical information available on the weight of spherical rubber centers. It was known that the process under statistical control yielded an average of 20.13 grams ($\bar{X}'$) and a standard deviation of .13 grams (σ). Thus, the three sigma limits just met the specification requirements of 20.13 $\pm$ 0.39 grams. Now the pertinent information for a "group weight control chart," given

Group Size = n = 5

Weight of Each Group Size = X_s, where s denotes sum.

Average Weight of X_s = $n\bar{X}'$ = $\bar{X}_s'$ = (5)(20.13) = 100.65 gms.

Standard Deviation of X_s = σ_s = $\sqrt{n}\sigma$ $\quad$ $(\sqrt{5})(.13)$ = .291 gms.

is

1) Average line

$$\bar{X}_s' = 100.65 \text{ gms}$$

2) Lower control Limit

$$\bar{X}_s' - 3\sigma_s = n\bar{X}' - 3\sqrt{n}\sigma = 99.78 \text{ gms.}$$

3) Upper Control Limit

$$\bar{X}_s' - 3\sigma_s = n\bar{X}' + 3\sqrt{n}\sigma = 101.52 \text{ gms.}$$

This information was immediately incorporated into a group weight control chart** and the data of Table VI plotted on this control chart showed a state of excellent statistical control for the two day production run.

---

* The increase in accuracy is obtained from minimizing balance errors. For example, a one-gram or a five-gram weight may each be weighed on a pan balance to the nearest $\pm$ .05 gram, with a possibility of 5 percent error in the first case and only 1 percent in the second.

** The group weight control chart is similar in all respects to the conventional $\bar{X}$ - chart except sums or totals are plotted instead of averages. The operating characteristic curves of the two charts are identical.

TABLE VI

DATE	TIME OF DAY	X_s - GROUP WEIGHT OF 5 INDIVIDUALS IN GRAMS
1-6	8 A.M.	100.9
"	11 A.M.	100.4
"	2 P.M.	100.8
1-7	8 A.M.	100.4
"	11 A.M.	100.5
"	2 P.M.	101.3

Because of the criticalness of the weight specification of the spherical rubber centers the QCE decided to treat group weights in the same way that individual determinations are used in the construction of $\bar{X}$ and R-Charts. He set up $\bar{X}_s$ and R_s charts in order to appraise the data of Table VII.

TABLE VII

DATE	TIME OF DAY	6 WEIGHINGS OF 5 INDIVIDUALS EACH (X_s)						$\bar{X}_s$	R_s
		1	2	3	4	5	6		
1-6	11 A.M.	100.7	100.2	100.7	100.3	100.7	101.4	100.67	1.2
1-6	2 P.M.	100.3	100.0	100.9	100.5	100.6	100.5	100.47	0.9
1-7	8 A.M.	99.9	100.4	101.0	101.2	100.6	100.6	100.47	1.3

1. $\bar{X}_s$ Chart

   a) Average Line

   $\bar{X}_s' = n\bar{X}' = 100.65$ gms.

   b) Lower Control Limit

   $\bar{X}_s' - 3\sigma_{\bar{X}_s} = \bar{X}_s' - \frac{3\sqrt{n}\sigma}{\sqrt{m}} = 100.65 - .36 = 100.29$ gms.

   where m is number of samples in each set

   and $\sigma_{\bar{X}_s} = \frac{\sqrt{n}\sigma}{\sqrt{m}} = \frac{(\sqrt{5})(.13)}{\sqrt{6}} = .12$ gms.

c) Upper Control Limit

$$\bar{\bar{X}}_s' + 3\sigma_{\bar{X}_s} = 100.65 + .36 = 101.01 \text{ gms.}$$

2. R_s Chart

With the tabular values of d_2, D_1 and D_2 of Table VIII we

TABLE VIII

NUMBER OF SAMPLES	d_2	D_1	D_2
2	1.128	0	3.69
3	1.693	0	4.36
4	2.059	0	4.70
5	2.326	0	4.92
6	2.534	0	5.08
7	2.704	0.20	5.20
8	2.847	0.39	5.31

may proceed as follows:

a) Average Line

$$\bar{R}_m' = d_{2,m}\sigma_s = d_2\sqrt{n}\,\sigma = (2.534)(\sqrt{5})(0.13) = 0.7 \text{ gms.}$$

b) Lower Control Limit

$$D_{1,m}\sigma_s = D_{1,m}\sqrt{n}\,\sigma = (0)(\sqrt{5})(0.13) = 0 \text{ gms.}$$

c) Upper Control Limit

$$D_{2,m}\sigma_s = D_{2,m}\sqrt{n}\,\sigma = (5.08)(\sqrt{5})(0.13) = 1.5 \text{ gms.}$$

When the above control values are used to construct the $\bar{\bar{X}}_s$ and R_s charts and when the actual values of $\bar{\bar{X}}_s$ and R_s are plotted (Table VII) on their respective charts we once again have a state of statistical control existing within the confines dictated by the specifications. Further the QCE knew that the $\bar{\bar{X}}_s$ chart was very powerful in detecting non-conformance parts. The extent of this power is contained in the formula*

$$d_{p(\bar{\bar{X}}_s)} = \sqrt{n \cdot m}\ (d_{X\%} + 3) - 3$$

* Formula assumes a normal distribution of X_s and a constant σ_s. The assumption of normality is highly realistic because sums are involved.

where $p(\bar{X}_s)$ is the probability of rejecting a process of X% non-conformance and where $d_{p(\bar{X}_s)}$ and $d_{X\%}$ are respectively the normal deviates. To illustrate, suppose the process is producing 1% of rubber centers below the lower specification of 19.74 grams. Then the probability that a sample average $\bar{X}_s$ - which is based on 6 groups and 5 within each group or 30 values in all - will be smaller than the lower control limit is

$$p(\bar{X}_s) = .75$$

since $n = 5$, $m = 6$, $X = 1\%$, $t_{X\%} = 2.33$ (Normal Table)

$$d_{p(\bar{X}_s)} = \sqrt{30}\,(-2.33 + 3) - 3 = -0.67 \text{ (Normal Table)}$$

Therefore, 75 times out of 100, the process will be rejected and call for an adjustment when it is producing 1% product below its lower limit. Or, in terms of its complement, the probability of accepting the process in one determination of $\bar{X}_s$ is .25. For two determinations, the probability of acceptance is .0625; for three it is .015625 and so forth.

The formula could also be used to calculate sample sizes for given AQL's and probability of acceptance (P_A). Suppose that

$$AQL = X = 1\% \qquad \therefore\ d_{X\%} = -2.33$$

$$P_A = 1 - p(\bar{X}_s) = 1\% \qquad \therefore\ d_{p(\bar{X}_s)} = +2.33$$

$$\sqrt{n.m} = \frac{d_{p(\bar{X}_s)} + 3}{d_{X\%} + 3} = \frac{5.33}{.67} = 8$$

$$\therefore\ n.m = 64$$

Therefore, one set satisfying this relation is $n = 8$ and $m = 8$. Another is $n = 4$, $m = 16$. The combination chosen is that one, which yields homogeneity conditions within each group.

All of the above discussion on rubber centers is based on known values of $\bar{X}'$ and σ. When these are unknown, then the conventional method is used to determine whether the process is in a state of statistical control. We obtain a number of sets (at least 10) and set up the R_s and $\bar{X}_s$ charts.

	$\bar{X}_s$ chart	R_s chart
a) Average Line	$\bar{\bar{X}}_s$	$\bar{R}_s$
b) Lower Control Limit	$\bar{\bar{X}}_s - A_2 \bar{R}_s$	$D_3 \bar{R}_s$
c) Upper Control Limit	$\bar{\bar{X}}_s + A_2 \bar{R}_s$	$D_4 \bar{R}_s$

where A_2, D_3 and D_4 are given in Table IX.

TABLE IX

Sample Size	2	3	4	5	6
A_2	1.88	1.02	0.73	0.58	0.48
D_3	0	0	0	0	0
D_4	3.27	2.57	2.28	2.11	2.00

Then, when both charts are in statistical control, due to corrective action having been taken in the process, we may calculate an estimate of the average and standard deviation of the individual weights of rubber centers. This may be done even though no individual reading was ever taken. These estimates would be obtained as follows:

$$S = \frac{\bar{R}_s}{d_{2,m}\sqrt{n}} \text{ and } \bar{X} = \frac{\bar{X}_s}{n}$$

These values are then used to determine whether specifications are met or what percent of product lies outside specification limits.

Several days later the QCE was given a problem which had the engineering department quite perplexed. Eight new flex life testing machines had been manufactured to a high degree of precision. Because of this everyone felt confident that each machine would test very nearly alike. However, it soon became apparent that in testing samples from a homogeneous batch there existed real machine differences. For the same product some machines would give high values and others low values. In large part, acceptance or rejection of lots depended on the testing machine chosen.

The task before the QCE was

1. To determine whether flex life measurements on two or more testing machines are equivalent to repeated measurements on the same machine. Each machine is compared to the acceptable standard machine, Y.

2. To provide a method of combining results when the testing machine differs from that of the standard (Y); that is, estimating Y for test results of, say, machine X.

As a vital step to this end, the QCE pointed out the importance of gathering proper data. In order to effectuate a comparison between two machines duplicate tests must be run on sample pairs which are as identical as possible - two samples from the same batch and adjacent cells in a mold. Also, pairs should each be taken from a different batch. In gathering data this way, it is logical to expect, apart from random variations and assuming the two machines X and Y are the same in measuring flex life, that the

1. differences, $d = X - Y$ and the mean of $\bar{d}$ to be zero

2. variances of Y and X (s_Y^2 and s_X^2) to be equal, and

3. equation of relationship to be $Y = X$, which signifies that $a = 0$ and $b = 1$ in the general equation of the straight line, $Y = a + bX$

In effect, then, the question of whether the two machines are alike is answered by statistically testing the realism of the above expectations.

The QCE proceeded to do this for 15 pairs of flex life measurements. Some of these values, along with the difference between X and Y, are given in Table X.

TABLE X

FLEX LIFE, IN THOUSANDS AND FURTHER CODED BY SUBTRACTING 350

Batch No.	Y	X	d = X - Y
1	- 15	- 76	91
2	- 39	- 36	75
3	- 18	- 149	131
.	.	.	.
.	.	.	.
.	.	.	.
14	- 59	- 23	82
15	- 37	- 45	82
	- 521	- 702	1223

The steps involved, in testing $\bar{d}$, in order to determine if it represents a real difference introduced by machines or a random difference of sampling, are:

1. Calculate the average difference $\bar{d}$

$$\bar{d} = \frac{1223}{15} = 81.9$$

2. Calculate the standard deviation of the differences, d.

$$S_d = 30.14$$

3. Determine

$$t = \frac{\sqrt{n}\,\bar{d}}{s_d} = \frac{(15)(81.9)}{30.14} = 10.52$$

4. Compare t and $t_{.05}$ for df = n - 1. The value of $t_{.05}$ for 14 df is 1.76. Therefore t is greater than $t_{.05}$. We conclude that machine X measures flex life by 81.9 units more on the average than machine Y.

The equality of variances of Y and X can be tested by comparing the variance ratio of

$$F = \frac{s_X^2}{s_Y^2} = \frac{5421.43}{2017.64} = 2.69$$

with $F_{.05}$ with 14 df for both numerator and denominator. $F_{.05}$ equals 2.48 and consequently we conclude the variation of machine X is greater than that of machine Y since F is greater than $F_{.05}$.

To answer the question concerning the equation of relationship the following steps are pertinent.

1. Calculate

$$b = \frac{s_Y^2 + s_X^2 - s_d^2}{2s_X^2} = \frac{2017.64 + 5421.43 - 908.41}{10842.86} = 0.602$$

2. Determine, statistically, if b is different from 1

$$t = \frac{\sqrt{n}(s_X)(1 - b)}{\sqrt{s_Y^2 - b^2 s_X^2}} = \frac{\sqrt{15}\sqrt{5421.43}\ (.398)}{\sqrt{2017.64-(.602)^2(5421.43)}} = \frac{113.5}{7.27} = 15.61$$

Compare t and $t_{.05}$ with df of n - 2. The value of $t_{.05}$ for 13 df is 1.77 and therefore the value of b cannot be taken as 1 but as definitely equal to .602.

3. Calculate

$$a = \bar{Y} - b\bar{X} = \frac{-521}{15} - \frac{(.602)(702)}{15} = -62.93$$

4. Determine, statistically, if a is different from zero

$$|t| = \frac{\sqrt{n}\, S_X\, a}{\sqrt{(S_Y^2 - b^2 S_X^2)(S_X^2 + \bar{X}^2)}} = \frac{(\sqrt{15})(73.63)(62.93)}{\sqrt{(52.89)(7611.67)}}$$

$$= \frac{17945.68}{634.44} = 28.29$$

This value of $|t|$ is greatly in excess of $t_{.05}$ for 13 df so that the conclusion is the a value cannot be taken as zero but is significantly equal to - 62.93 or - 63.

5. The equation of relationship for estimating the value of machine Y, given the test result of machine X is

$$Y = -63 + 0.6X$$

and in decoded terms, since $Y_D = Y - 350$ and $X_D = X - 350$, it is

$$Y_D = 77 + 0.6X_D$$

where the subscript D denotes decoded values. Thus, if flex life is 375,000 cycles on machine X then it is estimated that on machine Y the number of cycles would have been

$$Y_D = 77 + (.6)(375) = 302 \text{ or } 302{,}000 \text{ cycles}$$

6. The 90% confidence interval is given by

$$Y_D \pm \frac{t_{n-2}}{\sqrt{n}\, S_X} \sqrt{S_Y^2 - b^2 S_X^2} \sqrt{S_X^2 - (X_D - \bar{X}_D)^2}$$

$$302 \pm \frac{(1.77)(7.27)(120.02)}{(3.873)(73.63)} = 302 \pm \frac{1544.41}{285.169} = 302 \pm 5.4$$

Therefore the confidence interval is 296,600 to 307,400 cycles. When the part tested 375,000 cycles on machine X, there is a 90% confidence in the conclusion that the same part would have tested on the standard machine Y to some one value in the interval 296,600 to 307,400 cycles. The testing on machine X is not as severe as that of the standard machine Y.

After all the machines had been related to the standard by the above procedures the question arose as to which machine is the more closely related to the standard. A measure of closeness is given by

$$C = (100)(b^2)\left(\frac{S_X^2}{S_Y^2}\right) = (100)(.602)^2\ (2.69) = 97.49\%$$

This means that the machine with the highest value of C whose maximum is 100% is the most closely correlated to the standard machine Y. Or, 1 - C = 2.51%, gives the percent of error introduced by using machine X instead of machine Y. It is desirable to select those machines whose percent errors with respect to the standard machine of Y are the smallest.

Acceptance or rejection of lots no longer was dependent largely upon machine variations. The test results of each machine are translated statistically to that of the standard. This eliminates machine variations leaving product performance and variation. The engineering department was very pleased with the entire statistical approach.

The QCE's next encounter was with one of the salesmen. The salesman had just visited a customer who had given him a variable sampling plan that was to be applied to flex life for the acceptance or rejection of lots. The plan was simple.

1. Take a sample of ten from a lot and calculate $\bar{X} - 2.108S$.
2. If $\bar{X} - 2.108S$ is greater or equal to the lower specification, L, accept the lot otherwise reject it. This is written $\bar{X} - 2.108S \geq L$.

The salesman wanted to know how "stiff" was this plan? The QCE realized that this sampling plan was based upon the non-central t distribution since $\bar{X} - 2.108S \geq L$ is equivalent to $(L - \bar{X}) / S \leq -2.108$. Therefore, the "stiffness" of the plan was readily obtainable from Tables of the Non-Central t Distribution (3). The operation of the plan is given in Table XI. Thus, if flex life in a lot is normally distributed and 1% is in non-conformance to L, then the probability that the lot will be accepted by the sampling plan $\bar{X} - 2.108S \geq L$ is 82 out of 100, or 18 lots in 100 will be rejected. Even at 4/10 of 1%, 7 out of 100 lots would be rejected. The quality level of the customer was rather high.

Upon learning about this plan, the salesman decided to have the QCE review a sampling plan that a customer had sent in by mail. The customer wanted this plan to be applied to an upper specification of 230 lbs. for the 20% Dry Compression Test (Link Tester with specimens of $\frac{1}{2}$" by 1.129" cured for 15 minutes at 316° F. and then cooled in water for 10 minutes). The plan was as follows.

	Accept	Reject	Continue Sampling	
$n_1 = 5$	$C > 2.5$	$C \leq 1.5$	$1.5 < C < 2.5$	$C = \frac{U - \bar{X}}{S}$
$n_2 = 15$	-	-	-	and
$n = n_1 - n_2 = 20$	$C > 2.0$	$C \leq 2$	-	$U = 230$ lbs.

TABLE XI

LOT QUALITY (%)	PROBABILITY OF ACCEPTANCE P_A (%)
0.4	93
1.0	82
2.5	58
4.0	42
6.5	26
10.0	13
15.0	5
25.0	1

This plan says to take a first sample (n_1) of 5 and determine the value of C. If C is equal to or greater than 2.5 accept the batch for production; equal to or less than 1.0, reject the batch; and, between 1 and 2.5, obtain a second sample of 15. Then for the total of 20 sample results determine C and reject the batch if C is less than 1.5, otherwise, accept. The QCE decided to calculate by using tables of non-central t the chances of accepting (P_A) or rejecting (P_R) batches of quality 1, 2. 5 and 4.0% in non-conformance to the value of 230 lbs., assuming a normal distribution. The probabilities are:

Lot Quality	1%			2.5%			4.0%		
	P_A	P_R	Cont'd	P_A	P_R	Cont'd.	P_A	P_R	Cont'd
$n_1 = 5$	.47	.09	.44	.31	.22	.47	.23	.32	.45
$n = 20$	.36	.08	-	.23	.24	-	.13	.32	-
	.83	.17		.54	.46		.36	.64	

If batches are produced at 2.5% quality then the probability for the first sample of accepting is 31%, of rejecting 22%, and of continued sampling 47%. If continued sampling takes place then the probability for the combined sample of accepting is 23% and of rejecting 24%. The plan as a whole will accept 54% of the time and reject 46%. The salesman hadn't realized the complexity of the plan nor its severity. He certainly would talk to the production superintendent and the plant technologist before agreeing to accept it.

By this time the QCE and the salesman had become very good friends and consequently the salesman didn't hesitate to tell the QCE about a complaint from a very unhappy customer. The customer had complained

that many bushings of part number XY232 were quite eccentric and this was seriously interfering with the functioning of the finished unit. The salesman had discussed the complaint with the plant manager who said very little could be done because the variation in eccentricity from part to part was due to non-uniformity within the 100 cavity mold itself. In fact, the plant manager felt strongly that there were trend differences from column to column, row to row, and along any diagonal. In order to substantiate his feelings about this he had 100 bushings measured for concentricity and had each measurement recorded in Table XII in the similar position the bushing was formed in the mold - for convenience the columns and rows are symbolized by X's and Y's respectively. However, after the table was completely filled in, the plant manager noticed a

TABLE XII

CODED VALUES OF CONCENTRICITY OF XY232

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	$\bar{Y}$
Y_1	5	2	6	5	4	4	4	4	4	6	4.4
Y_2	5	3	4	2	3	2	4	4	1	4	3.2
Y_3	4	2	2	2	2	3	3	2	1	4	2.5
Y_4	2	2	2	2	3	2	2	2	2	4	2.3
Y_5	2	2	2	2	3	2	2	4	3	2	2.4
Y_6	3	0	2	4	4	2	3	3	2	4	2.7
Y_7	2	1	4	1	4	4	4	4	2	0	2.6
Y_8	2	4	3	2	3	3	4	4	4	4	3.3
Y_9	3	4	3	4	3	4	5	4	4	2	3.6
Y_{10}	6	6	6	4	4	6	6	6	6	4	5.4
$\bar{X}$	3.4	2.6	3.4	2.8	3.3	3.2	3.7	3.7	2.9	3.4	

"trend" variation among the averages of the rows and also noticed quite a "hilly" effect in the averages of the columns. This information was obtained by observation only. The salesman wanted to know what could be done by statistical analysis to ferret out the maximum amount of information from the data.

The QCE decided to employ the Analysis of Variance technique to test statistically whether there were any significant linear, quadratic, cubic, quartic, and interaction effects. As a first step, a simple two factor analysis of the data gave the results in Table XIII.* A comparison of the

* Computational steps for this problem are omitted

TABLE XIII

SOURCE OF VARIANCE	SUM OF SQUARES	df	MEAN SQUARE	F-RATIOS	$F_{.05}$	
Rows (Y)	89.84	9	9.98	9.15	1.99	S
Columns (X)	12.24	9	1.36	1.25	1.99	NS
Residual	88.16	81	1.09			
Total	190.24	99				

TABLE XIV

SOURCE OF VARIANCE		SUM OF SQUARES	df	MEAN SQUARE	F-RATIOS	$F_{.05}$	TEST
Y: Row							
Linear	(L)	8.76	1	8.76	8.05	3.96	S
Quadratic	(Q)	75.17	1	75.17	69.06	3.96	S
Cubic	(C)	0.02	1	0.02	54.50	252.00	NS
Quartic	(q)	4.21	1	4.21	3.87	3.96	NS
Remainder		1.68	5	0.34	32.06	252.00	NS
X: Column							
Linear	(L)	1.16	1	1.16	1.06	3.96	NS
Quadratic	(Q)	0.05	1	0.05	21.80	252.00	NS
Cubic	(C)	2.00	1	2.00	1.84	3.96	NS
Quartic	(q)	1.03	1	1.03	1.06	252.00	NS
Remainder		8.00	5	1.60	1.47	2.33	NS
Residual		88.16	81	1.09	-		

Note: All sixteen interactions $X_L Y_L$, $X_L Y_Q$ . . . $X_q Y_q$ tested non-significant and thus were pooled with residual.

TABLE XV

SOURCE OF VARIANCE		SUM OF SQUARES	df	MEAN SQUARE	F-RATIOS	$F_{.05}$	TEST
Y: Row							
Linear	(L)	8.76	1	8.76	7.99	3.94	S
Quadratic	(Q)	75.17	1	75.17	68.59	3.94	S
Residual		106.31	97	1.10	--	--	

F-ratios with $F_{.05}$ shows a high significance for rows and non-significance for columns. The Analysis of Variance for trend effects using the Method of Orthogonal Polynomials (4) yields the results contained in Tables XIV and XV. The analysis clearly indicates that a quadratic effect in the rows is very significant. Consequently, the QCE proceeded to fit a quadratic equation by the Method of Least Squares to the averages of the rows and obtained

$$\bar{Y}_E = 2.26 + 0.052Y + 0.03Y^2$$

where $Y_1 = -9$, $Y_2 = -7$, . . . , $Y_9 = +7$, $Y_{10} = +9$ and $\bar{Y}_E$ is the expected average of each row. A comparison of actual to expected is presented in Table XVI. The tenth row of the mold gives the highest expected average

TABLE XVI

	Actual	Expected
Y_1	4.4	4.21
Y_2	3.2	3.36
Y_3	2.5	2.74
Y_4	2.3	2.37
Y_5	2.4	2.23
Y_6	2.7	2.34
Y_7	2.6	2.68
Y_8	3.3	3.26
Y_9	3.6	4.08
Y_{10}	5.4	5.13

with an individual variation as high as

$$5.13 + (1.96)\sqrt{1.10} = 5.13 + 2.06 = 7.19$$

The lowest average is in row five and the individual variation is as low as

$$2.23 - (1.96)\sqrt{1.10} = 2.23 - 2.06 = 0.17$$

The overall variation was therefore 7.02 units (coded) of which the mold itself contributes 41.3%. With the elimination of the quadratic mold effect and assuming that the smallest expected average of 2.23 would characterize each row, then the highest value would be

$$2.23 + (1.96)\sqrt{1.10} = 2.23 + 2.06 = 2.29$$

a value which would please the customer greatly. The QCE, salesman and plant manager agreed that the mold contribution had to be eliminated and both were very pleased with the statistical analysis.

The QCE's next important consideration was the process effectiveness of batch blending. A number, B, of initial banbury batches are blended into an intermediate batch; then a number, I, of intermediate batches are blended into a finished batch, and a number, f, of finished batches are blended into a final batch. A schematic representation of this is given in Table XVII.

Since the effect of perfect blending is identical to averaging, it follows that the intermediate batch is the average of the banbury batches. In Table XVII, therefore, the I_1 -value of 226 is the average of the four values B_1 = 220, B_2 = 214, B_3 = 240 and B_4 = 230. Similarly the f_1-value of 223 is the average of I_1 = 226, I_2 = 219, I_3 = 222 and I_4 = 225; the f_1-value of 224 is the average of f_1 = 223, f_2 = 220 and f_3 = 229. It also follows that blending produces more uniform batches since blending or averaging eliminates variation. This variation is of two kinds:

1) the variation of the heterogeneous nature of raw rubber, and

2) the sampling variation of blending (or averaging) which is related to the number of batches blended.

As to the banbury batches, the within variation is due to the random variation of the raw rubber. Among the intermediate batches the total variation is composed of a portion of the within variation of the banbury batches and the possibility of an intermediate batch to batch raw rubber variation. Likewise, the total variation for the finished batches is composed of a portion of the within variation of the intermediate batches and the possibility of a finished batch to batch raw rubber variation. And, the total variation for the final batches is made up of a portion of the within variation of the finished batches and a possibility of a final batch to batch raw rubber variation.

Therefore an important question is how much uniformity is obtained with each successive step in blending. The QCE outlined the steps necessary to answer this question.

1. Determine the total within pooled variances of

   a) Banbury batches = S_B^2 = 400

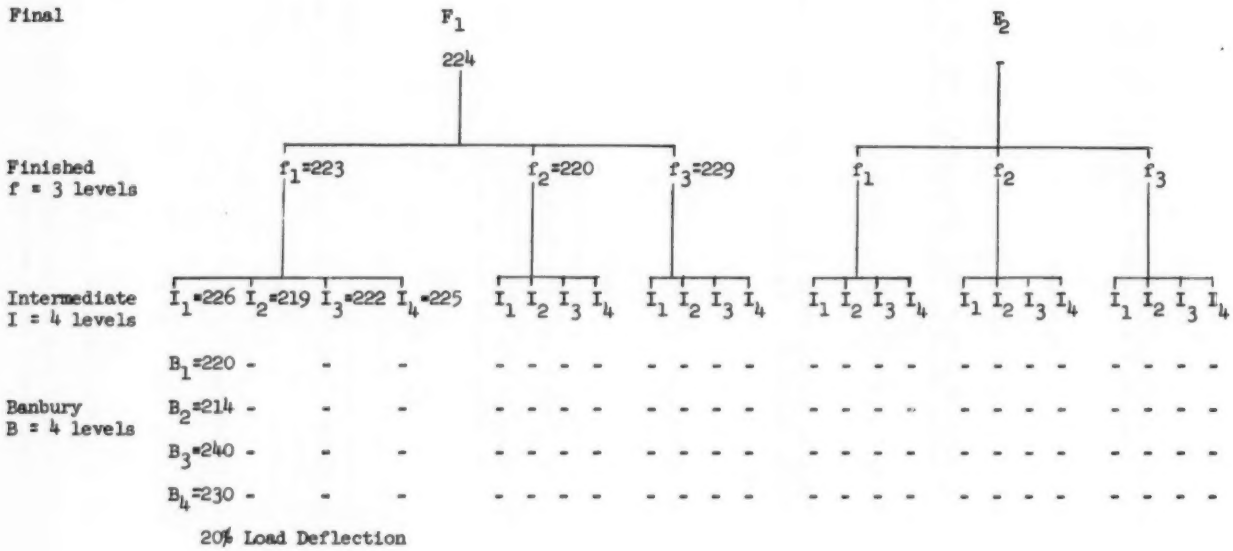
   b) Intermediate batches = S_I^2 = 180

   c) Finished batches = S_f^2 = 135

   d) Final batches = S_F^2 = 90

2. Conclude whether there is an intermediate batch to batch raw rubber variation - symbolized by S_I^2.

TABLE XVII



$S_B^2 = 400$ $S_I^2 = 180$ $S_f^2 = 135$ $S_F^2 = 90$

a) Determine $F = \dfrac{\dfrac{Bs_I^2}{(18df)}}{\dfrac{s_B^2}{(72df)}} = \dfrac{(4)(180)}{400} = 1.8$

b) $F_{.05}$ for 18 and 72 df is 1.75

c) Conclude that a real intermediate batch to batch raw rubber variation exists with a variance value of

$$s_{I'}^2 = s_I^2 - \frac{s_B^2}{B} = 180 - \frac{400}{4} = 80$$

3. Conclude whether there is a finished batch to batch raw rubber variation - symbolized by $s_{f'}^2$.

a) Determine $F = \dfrac{\dfrac{I\, s_f^2}{(4df)}}{\dfrac{s_I^2}{(18df)}} = \dfrac{(4)(135)}{180} = 3$

b) $F_{.05}$ for 4 and 18 df is 2.93

c) Conclude that a real finished batch to batch raw rubber variation exists with a variance value of

$$s_{f'}^2 = s_f^2 - \frac{s_I^2}{I} = 135 - \frac{180}{4} = 90$$

4. Conclude whether there is a final batch to batch raw rubber variation - symbolized by $s_{F'}^2$.

a) Determine $F = \dfrac{\dfrac{f\, s_F^2}{(1\ df)}}{\dfrac{s_f}{(4\ df)}} = \dfrac{(3)(90)}{135} = \dfrac{270}{135} = 2$

b) $F_{.05}$ for 1 and 4 df is 7.71

c) Conclude that a final batch to batch raw rubber variation doesn't exist and thus

$$S^2_{F'} = 0$$

5. Determine the total variance of rubber variation by adding S^2_B, $S^2_{I'}$, and $S^2_{f'}$.

$$S^2_T = S^2_B + S^2_{I'} + S^2_{f'} = 400 + 80 + 90 = 570$$

6. Uniformity Evaluations

a) Intermediate Batches: The uniformity of this blend is

$$U_I = 1 - \frac{S^2_I}{S^2_T} = 1 - \frac{180}{570} = 1 - 0.316 = 68.4\%$$

The first blend gives a uniformity value of 68.4%.

b) Finished batches: The uniformity of this blend is

$$U_f = 1 - \frac{S^2_f}{S^2_T} = 1 - \frac{135}{570} = 1 - 0.237 = 76.3\%$$

Thus, the second blend has increased the uniformity of the batches by 73.6 - 68.4 or 5.2%.

c) Final Batches: The uniformity of this blend is

$$U_F = 1 - \frac{90}{570} = 1 - 0.158 = 84.2\%$$

The third blend has increased the uniformity by 84.2 - 76.3 or 7.9%.

After the above calculations and conclusions the QCE decided to determine how much uniformity would be obtained by blending the final batches in groups of four - assuming this to be economically and physically possible. The equation used for this purpose is

$$S^2 = \frac{S^2_B}{BIfF} + \frac{S^2_{I'}}{IfF} + \frac{S^2_{f'}}{fF}$$

where B = 4, I = 4, f = 3 and F = 4 and

$$U = 1 - \frac{S^2}{S^2_T} = 1 - 0.0197 = 98.03\%$$

Almost perfect uniformity would be obtained. However, the more practical approach to the problem would be to purchase more uniform rubber so as to eliminate S^2_I, and S^2_r, and reduce the value of S^2_B. This would cut out the need for successive blending operations. The QCE also realized that perfect blending is an impossibility so that this also called out for uniform rubber. Imperfect blending could be a strong factor in grading downward the above uniformity evaluations.

The QCE and one of his quality control cohorts were talking about the natural tendency of flex life to be positively skewed. The QCE had a feeling this followed because theoretically the minimum value is zero and the maximum is infinity. Or better yet, rubber compound is formulated to yield flex life values which will safely meet the minimum requirement. As a result flex life values will tend to concentrate at some one value above but near the minimum requirement and then to stretch out with a gradual decrease in frequency into higher regions of long life. Also the QCE indicated that a reciprocal transformation might transform the distribution of flex life into a normal distribution because by taking reciprocals the variation between smaller values is enlarged where as between larger numbers it is decreased.

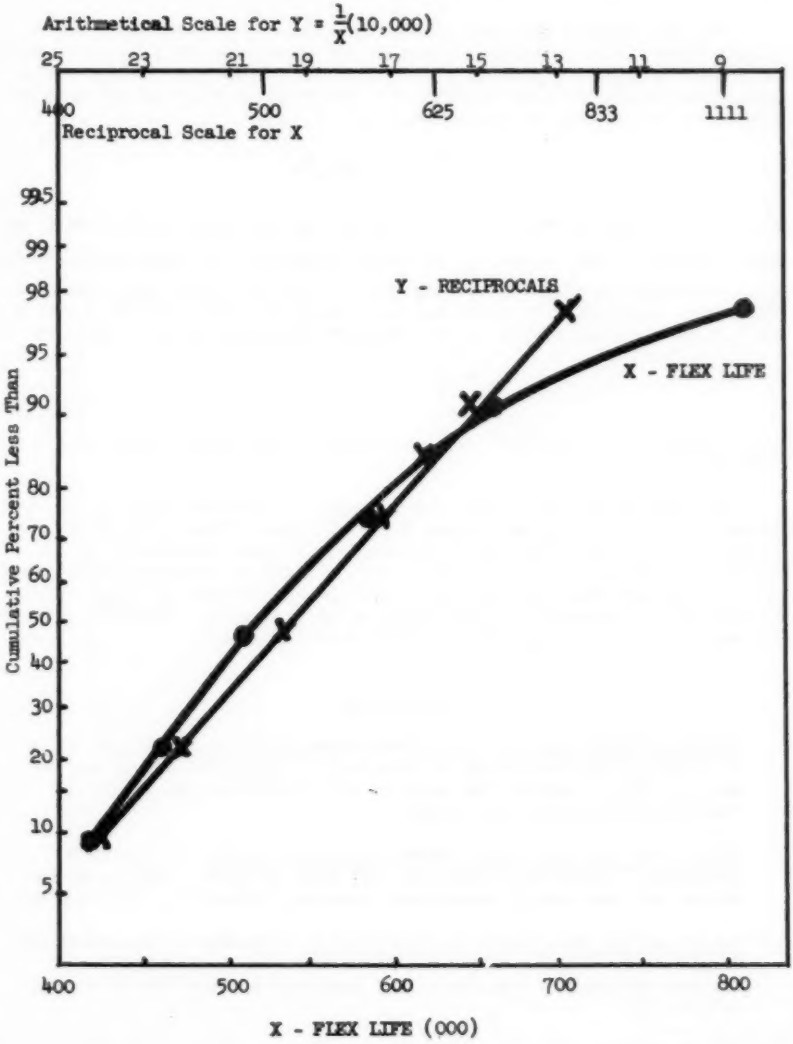
In order to substantiate this theorizing the QCE requested a series of flex life values from a homogeneous batch and was given the flex values of Table XVIII. This table also presents the reciprocal values and the cumulative less-than percent distribution. These percents have been adjusted for small sample size (5).

TABLE XVIII

X	$Y = (\frac{1}{X})(10{,}000)$	
FLEX LIFE	RECIPROCALS OF FLEX LIFE	CUMULATIVE LESS - THAN (%)
343	29.15	-
415	24.10	9.8
424	23.58	-
454	22.03	21.6
462	21.64	-
486	20.58	-
490	20.41	-
507	19.72	44.4
562	17.79	-
563	17.76	-
564	17.73	-
574	17.42	-
575	17.39	72.6
600	16.67	-
620	16.13	84.2
657	15.22	90.2
806	12.41	97.5

CHART 1

NORMAL PROBABILITY PAPER



When the cumulative percents are plotted on normal probability paper as in Chart 1, we note that flex values plot as a curve whereas the reciprocals plot as a straight line. This shows that the skewness of flex values can be eliminated by using reciprocal values. Further the transformation leads to a normal distribution. Thus, the above theorizing is substantiated by this set of values.

The QCE pointed out that it wasn't necessary to determine the reciprocals because the arithmetic scale of the normal probability paper could be scaled to plot reciprocals when the actual values are plotted. The reciprocal scale is shown in Chart 1. The formula relating the mean of flex life to the mean of reciprocals is

$$\bar{Y} = \frac{1}{\bar{X}} (1 - r_{XY} S_X S_Y)$$

where r_{XY} is the coefficient of variation and the other symbols have the usual meaning. Now, since r_{XY} is small because of the equilateral hyperbola relationship of X to Y we may use $\bar{Y} = 1/\bar{X}$ as a good approximation. The actual value of $\bar{Y}$ = .00194 and by formula we have 1/535.41 = .00187. Also, a good approximation to the standard deviation of Y is given by

$$S_Y = \frac{S_X}{\bar{X}^2}$$

which value is $104.581/(535.41)^2$ or .000365. The actual value is .000383.

The QCE wanted to discuss other topics of interest with his colleague. For instance, the actual use of Binomial Probability Paper, Statistical Process Controls, Experimental Designs, Tolerances by the Statistical Method, and Attribute Sampling Plans to problems in assembly, curing time and temperatures, gage sensitivity, control of flash and acceptance or rejection of lots would be entertaining. However, since the day was at a close he decided to defer this discussion.

References

1. Sampling Tables for Variables Inspection Based on the Range, by W. Grant Ireson and George J. Resnikoff. Technical Report No. 11, Nov. 5, 1952. Applied Mathematics and Statistics Laboratory, Stanford University, California.

2. Tables for One-Sided Statistical Tolerance Limits, by Gerald J. Lieberman. Technical Report No. 34, Nov. 1, 1957. Applied Mathematics and Statistics Laboratory, Stanford University, California.

3. Tables of the Non-Central t-Distribution by George J. Resnikoff and Gerald J. Lieberman. Technical Report No. 32, April 1, 1957. Applied Mathematics and Statistics Laboratory, Stanford University, California.

4. Design and Analysis of Industrial Experiments, by O. L. Davies. Hafner Publishing Company, New York, 1954. Chapter 8.

5. Sampling Inspection by Variables with no Calculations, by Herman Chernoff and Gerald J. Lieberman. Technical Report No. 22, April 22, 1955. Applied Mathematics and Statistics Laboratory, Stanford University, California.

AN ORGANIZED APPROACH TO RELIABILITY

Erroll J. Lancaster
Ballistic Missile Center, Air Materiel Command

During the past decade, reliability has become a factor of major significance in the development, production and operation of military equipment. The high operational reliability requirements of modern weapons and the high cost of attaining and maintaining this reliability have focused attention on the need for an organized approach to reliability assurance. The purpose of this paper is to suggest such an approach for application to the development and production of complex weapons.

Most of you, who are even casually familiar with Government contracting recognize that there has been a continually increasing emphasis on reliability in recent years, but to date there has been no uniform approach to establishing and enforcing reliability requirements in contracts. Some Government contracts simply refer to the need for considering reliability as a design parameter while others actually express quantitative reliability values. At best, however, there is considerable ambiguity as to what actually is required of the contractor. This ambiguity leads to difficulty on the part of contractors in estimating the cost of reliability when preparing competitive proposals. It also leads to difficulty on the part of the Government in evaluating the adequacy of contractor proposals for award of contracts and in determining whether, in fact, the contractor has fulfilled the reliability requirements of the contract. Lack of a uniform approach can only lead to inequities to both the Government and industry.

It seems obvious, therefore, that a general reliability specification, acceptable to industry and the Government, is fundamental to establishing a clear understanding of what is desired. Such a document should establish the basic reliability concepts and a systematic approach to the problem of assuring reliability.

I think you will all agree that reliability, like other qualitative factors, cannot be inspected or tested into an item. It can only result from a systematic process of control for each and every step from the design through the operational phase. The Government, therefore, to obtain reasonable assurance of attaining reliability should "buy" the contractors reliability system as well as the end product. That is, the Government should define the minimum elements of an acceptable reliability system that the contractor would legally be required to establish. Most of you will recognize this approach as basically the same one upon which the general Air Force quality control specification is based, and actually the problems are quite similar in nature although the basic elements differ in many respects.

From a responsibility point of view, it appears obvious to me that the prime contractor must be assigned the basic responsibility for reliability, including adequate assurance of reliability in subcontracted items. The Government in its role as a consumer, has the responsibility for monitoring or exercising surveillance over the contractors reliability efforts, and I believe, should play a more active role in this regard. I will discuss this subject of surveillance in more detail

later, but first let's consider some of the elements of a general reliability specification.

I believe one of the most important elements of the specification would be the definition of the various terms associated with the subject of reliability. The inclusion of these definitions in a general reliability specification would fulfill a current need for standardization of terminology in a media which would be recognized and accepted. While considerable effort has been expended on the definition problem and general agreements have been reached in certain segments of industry and Government, there is no recognized publication of reliability terminology which is accepted throughout the government-industry complex. I believe most of you will agree that semantics is one of the basic problems we face in this area of reliability.

Closely akin to the problem of definition of terms is the problem of defining the basic reliability concepts or principles. A clear statement of these fundamental principles is essential in order that we can all look at reliability through the same basic frame of reference.

As I mentioned previously, reliability can only result from a systematic process of control. The planning function is therefore a factor of major significance and the specification should clearly reflect this. Essentially, it should require that the contractor prepare a written description of his reliability program and the procedures to implement this program. The program should include a time-phased plan for attaining the quantitative reliability requirements contained in the contract and the manner in which the contractor will demonstrate attainment of these requirements.

In addition to fulfilling a basic need of the contractor, the written program and procedures represent a fundamental element in conducting surveillance on the part of the Government and would promote better understanding between the Government and the contractor.

The primary function of the specification, of course, would be to establish the minimum acceptable elements of a reliability program. It would be impossible in a paper of this length to discuss all of the elements required so I will suggest some elements as examples of what should be included in this part of the specification.

The specification should require preliminary and continuing studies to establish as early as possible and to revise, as necessary, quantitative reliability goals for systems, sub-systems, and functional performance limits and the environmental conditions under which the goals are to be achieved. These studies should cover all phases, from design goals through fabrication, testing, transportation, storage, maintenance and flight. They should determine the possibilities of trade offs of such parameters as weight, volume, performance, safety, storage life and producability versus reliability.

The specification should require that the contractors have a planned and scheduled program for the functional testing of a sample of production units under environmental conditions to measure the achieved reliability. In this regard, all equipment items, determined by the

reliability studies to have a significant bearing on reliability should be subjected to these tests.

Another example would be the requirement for the contractor to establish a system for failure reporting and analysis. This system should cover all aspects of the contractor's operations and should provide for rapid feedback of information and failed parts for analysis as required. The contractor should identify a single organization or activity responsible for analysis of failure data with proper authority to assure timely corrective action.

These three examples indicate, very briefly, what I have in mind as elements of a reliability program, other elements would include such things as, design testing, design review and evaluation, acceptance criteria, quality control, qualification testing, supplier or sub-contractor control, etc.

Considering the nature of the suggested specification, it is quite apparent that a surveillance program similar in nature to the one now exercised over the contractor's quality control system would be appropriate. Such a surveillance program would require highly qualified professional engineers with suitable training in statistical concepts and methodology.

In conclusion, I wish to again stress the importance of establishing a common base or standard which will serve as a suitable foundation for an organized approach to weapon system reliability. All of us either in Government or industry share a responsibility in the fulfillment of this objective.

QUALITY CONTROL CHARTS BY GAUGING

Leo A. Aroian
Systems Development Laboratories
Hughes Aircraft Company

1. Introduction. Since the inception of quality control charts by Dr. Shewhart, process control has been efficiently maintained and processes improved by the quality control chart for the average of the process, and the quality control chart for the variability of the process, employing the range. Economical process control is the goal of each of us as it has always been with Dr. Shewhart. Gauging is the answer to the expenses of exact measurement and the detailed arithmetic involved in the usual $(\bar{x},R)$ charting. This paper advocates a more widespread use of control charts based on gauging. Sample sizes and operating characteristic curves are provided.

2. Fundamental principles. Let x_1 be the lower gauge limit. The first gauge will be used to classify all items less than or equal to x_1. The second gauge will be used to classify all items greater than or equal to x_2, the upper gauge limit. The middle class will contain all items between x_1 and x_2. Consider the 3 frequency functions A, B, and C, with the gauge limits x_1 and x_2.

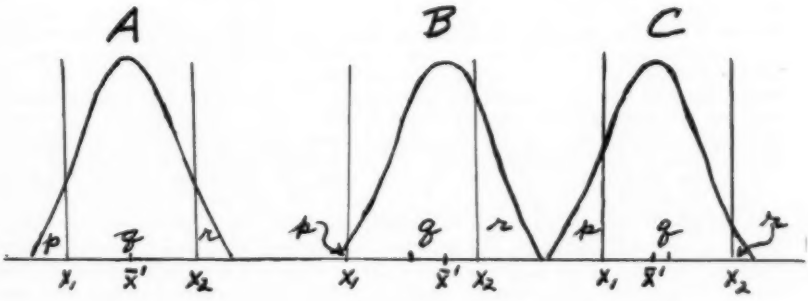


Figure I.

As $\bar{x}'$, the population mean increases, with no other changes in the frequency function, we note that r, the proportion of items falling above x_2, increases, while p, the proportion of items falling below x_1, decreases. Compare B with A. We define $q = 1 - p - r$, or q is the probability that an item will fall between x_1 and x_2. As $\bar{x}$ decreases, r decreases and p increases. Compare C with A. Hence r - p may be taken for control purposes for the mean. In a sample of n, we represent by c the number of items above x_2, the upper gauge limit, and by a the number of items below x_1, the lower gauge limit.

Similarly from Figure II, we see that a frequency function with a narrow spread stays within the gauge limits and a frequency function with a widespread spills over the gauge limits. Compare A, B, and C.

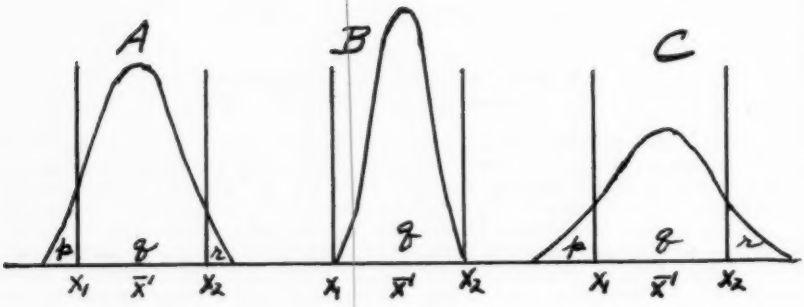


Figure II.

This suggests that r + p in the population is an indication of variability and that in the sample c + a is a measure of variability. We shall use c - a as the indicator for the mean chart and c + a as the indicator for the variability chart. We note that an increase in the mean will affect the variability chart in this system, but correspondingly in the usual $(\bar{x},R)$ control system a change in the variability will affect the mean chart. In neither case will there be difficulty of interpretation when both charts are used together.

3. Control limits for the mean. The distribution of c - a is found by use of the multinomial $(p + q + r)^n$. The mean of c - a is $n(r - p)$ and its variance is $n\{(p + r) - (p - r)^2\}$. Clearly we must choose p = r for control of the mean since the normal frequency function is symmetrical. Stevens (1) in his illuminating investigation of the best values of p, q, and r has used efficiency as his criterion. We shall instead use the operating characteristic curve as our guide, thus closing any theoretical gap existing in the theory of gauging. For the frequency function we choose the normal curve with mean $\bar{x}$ and standard deviation σ'. When the processes are in control we assume $\bar{x} = 0$ and $\sigma = 1$. We take n = 8, 10, and 12, and evaluate the distribution function of c - a, where c - a goes from n to -n in all cases. In Table I we list the values of x_1 and x_2. These remain the same for a particular choice of p, q, and r. In Table III we list the values of c - a for the control limits. We also list α, the probability that c - a will be greater than the upper control limit when the process is in control with mean zero and variance 1. The probability that c - a will exceed the upper control limit or will be less than the lower control limit will of course be 2α, when the process is on control. In the determination of the control limits we have chosen α such that $.001 = \alpha = .025$. To determine the best choice of p, q, and r we compare the O.C. curve (operating characteristic curve) for the c - a chart with the O.C. curve of the mean, adjusting both to have the same value $1 - \alpha$ at $\bar{x} = 0$. Actually we have calculated the power function of both charts which is merely one minus the ordinate of the O.C. curve. Stevens, by use of the criterion of efficiency, has indicated that p =.27, q =.46, r =.27 is the best choice with an efficiency of 81 . For p = r =.40, q =.20, the efficiency is 75 and for p = r =.20, q =.60, the efficiency has fallen only to 78 . In this region the change in efficiency is very slow. We have therefore made our investigations in the region starting with p = r =.40, q =.20, and ending with p = r =.05, q =.90, at intervals of .05 for p. The most efficient set of charts would be the $(\bar{x},\sigma)$ set but since this entails too much cost

practically everyone uses the $(\bar{x},R)$ set where R indicates the sample range. For the power function of the mean chart we have chosen sample sizes which will give the same power function as the samples of 8, 10, and 12 for the c - a chart. We then choose that value of p, q, and r which will give the best power function for c - a when compared against the corresponding power function of the sample mean. In all cases we have used the exact distribution of c - a which was coded on the IBM 704.

Table I.
Gauge Limits

p	q	r	x_1	x_2
.40	.20	.40	-.2533	.2533
.35	.30	.35	-.3854	.3854
.30	.40	.30	-.5244	.5244
.25	.50	.25	-.6745	.6745
.20	.60	.20	-.8416	.8416
.15	.70	.15	-1.036	1.036
.10	.80	.10	-1.282	1.282
.05	.90	.05	-1.645	1.645
.01	.98	.01	-2.326	2.326

4. Discussion of results for the mean chart. In columns 1, 2, and 3 of Table III are listed p, q, and r. In column 4 we have α, in column 5 the lower control limit for c - a, and in column 6 the upper control limit for c - a. In column 7 we have the equivalent n for the $\bar{x}$ chart which gives roughly the same O.C. curve as the c - a chart. In some cases the n is a fraction, but since this cannot occur we have used the n which gives the closest O.C. curve. We have indicated by a star the recommended sets when using the mean chart alone and by a dagger when using both charts together. For n = 12, the power efficiency is only 75 which occurs for p = r =.25, q =.50, while Stevens quotes the figure 80.6 . Of course, the method of Stevens is theoretically incorrect since he has based his theory on estimation of the parameters rather than on the distribution of the statistics. Further, only certain values of α are available. For example, α = .0054, n = 12, p = r =.20, q =.60 has a much better O.C. curve than α = .0033, n = 12, p = r =.25, q =.50 . The two power functions are given in Table II.

Table II.

p	q	r	$\bar{x}'$					
			0	.5	1.0	1.5	2.0	2.5
.25	.50	.25	.0033	.115	.610	.956	.999	1.000
.20	.60	.20	.0054	.152	.681	.974	1.000	1.000

In this case p = r =.20, q =.60 is the preferred set.

Table III.

n = 12

p	q	r	α	c - a		n
.40	.20	.40	.0022	-8.5	8.5	8
			.0066	-7.5	7.5	
			.0167	-6.5	6.5	
.35	.30	.35	.0011	-8.5	8.5	8
			.0039	-7.5	7.5	
			.0112	-6.5	6.5	
*.30	.40	.30	.0019	-7.5	8.5	9
*			.0067	-6.5	7.5	
*			.0190	-5.5	6.5	
*.25	.50	.25	.0033	-6.5	6.5	9
*			.0113	-5.5	5.5	
*†.20	.60	.20	.0012	-6.5	6.5	9
*†			.0054	-5.5	5.5	
*†			.0189	-4.5	4.5	
*†.15	.70	.15	.0018	-5.5	5.5	8
*†			.0084	-4.5	4.5	
*†.10	.80	.10	.0021	-5.5	5.5	8
*†			.0119	-4.5	4.5	
†.05	.90	.05	.0015	-4.5	4.5	7
			.0128	-3.5	3.5	

Table III. (cont.)

n = 10

p	q	r	α	c - a		n
.40	.20	.40	.0029	-7.5	7.5	6
			.0091	-6.5	6.5	
			.0025	-5.5	5.5	
.35	.30	.35	.0015	-7.5	7.5	7
			.0057	-6.5	6.5	
			.0173	-5.5	5.5	
*.30	.40	.30	.0030	-6.5	6.5	7
*			.0110	-5.5	5.5	8
.25	.50	.25	.0013	-6.5	6.5	8
*			.0059	-5.5	5.5	
*			.0207	-4.5	4.5	
*.20	.60	.20	.0025	-5.5	5.5	8
*			.0112	-4.5	4.5	
†.15	.70	.15	.0044	-4.5	4.5	7
†			.0205	-3.5	3.5	
†.10	.80	.10	.0070	-3.5	3.5	7
†.05	.90	.05	.0082	-2.5	2.5	6

n = 8

p	q	r	α	c - a		n
.40	.20	.40	.0033	-6.5	6.5	5
			.0131	-5.5	5.5	
.35	.30	.35	.0018	-6.5	6.5	5
			.0082	-5.5	5.5	6
.30	.40	.30	.0046	-5.5	5.5	6
			.0182	-4.5	4.5	
*.25	.50	.25	.0021	-5.5	5.5	6
*			.0106	-4.5	4.5	
*.20	.60	.20	.0050	-4.5	4.5	6
*			.0235	-3.5	3.5	
*†.10	.80	.10	.0033	-3.5	3.5	6
*†			.0231	-2.5	2.5	
*†.05	.90	.05	.0045	-2.5	2.5	5

A star indicates that the results are recommended if the mean chart is used alone.

A dagger indicates that the results are recommended if both the chart for control of the mean and the chart for control of the variability are used together.

5. Control limits for variability. In Table IV we list the control limits for c + a which ranges from o to n, with the same values of n as before. The distribution of c + a is a binomial distribution with mean $n(r + p)$ and variance $nq(r + p)$. We indicate by α_1 the probability that c + a will be less than the lower control limit and by α_2 the probability that c + a will exceed the upper control limit. In this case the upper control limit and the lower control limit will not be symmetrically placed. We have placed the limitation on α_1 as $.001 \leq \alpha_1 \leq .025$, and the limitation on α_2 of $.001 \leq \alpha_2 \leq .025$, except in a few cases. The power function of the c + a chart is compared with the power function of the range chart, and that set of p, q, and r will be judged best that maximizes the ratio of the power function of the c + a chart as compared with the power function of the range chart. Naturally the values of α_1 and α_2 for the two charts are chosen to be exactly the same. Stevens has shown that the maximum efficiency of the c + a chart is 65% with $p = r = .07$, $q = .86$. We have picked the values of n for the sample sizes of the range chart in such a way that the range chart and the c + a chart will have the same O.C. curves. Table IV in column 4 lists α_1 ; in column 5, the lower control limit for c + a; in column 6, α_2 ; in column 7, the upper control limit for c + a; in column 8 the equivalent value of n for the O.C. curve of the range (lower limit only); and in column 9 the equivalent value of n for the O.C. curve of the range (the upper limit only). We have not bothered to find the optimum value of p, q, and r. For n =12 we find that $p = .01 = r$, $q = .98$ is definitely better than $p = .05 = r$, $q = .90$. This differs from Stevens' results, but our comparison shows for $p = .05 = r$, $q = .90$, a power efficiency of 83 1/3% as compared with his figure of 64%. However, his comparison is with the standard deviation, while our comparison is against the range. Usually we wish to guard against an increase in variability, so the upper control limit should be used. However, we have given two-sided limits in case we wish to guard against either a decreases or an increase in variability. The dashes in Table IV indicate that no limits exist for those values of p, q, and r. The distribution function of c + a we obtain from (2). The distribution of the range was obtained from the Biometrika Tables for Statisticians (3).

6. Choice of joint control limits. Since the optimum choice of p, q, and r for the c - a chart is not the same as the optimum choice for the c + a chart, we must make some compromise. The c - a and c + a charts are not independent. We offer a solution based on the examination of the separate power functions. The optimum solution would involve the joint power function of c - a and c + a which is much too lengthy a task. The suggested choices for the joint control limits are indicated by a dagger in Tables III and IV.

Table IV.

n = 12

p	q	r	α_1	c + a	α_2	c + a		n
.20	.60	.20	.0022	.5	.0028	9.5	12	3
*†			.020	1.5	.015	8.5	12	4
*†.15	.70	.15	-	-	.0017	8.5	-	4
*†			.014	.5	.0095	7.5	10	5
*†.10	.80	.10	-	-	.0039	6.5	-	6
†			-	-	.019	5.5	-	7
*†.05	.90	.05	-	-	.0043	4.5	-	8
* .01.	.98	.01	-	-	.0015	2.5	-	10
			-	-	.023	1.5	-	10

n = 10

p	q	r	α_1	c + a	α_2	c + a		n
.15	.70	.15	-	-	.0016	7.5	-	3
*†			.028	.5	.011	6.5	10	4
*†.10	.80	.10	-	-	.0064	5.5	-	5
*†			-	-	.033	4.5	-	6
*†.05	.90	.05	-	-	.0016	4.5	-	7
*†			-	-	.013	3.5	-	7

n = 8

p	q	r	α_1	c + a	α_2	c + a		n
.10	.80	.10	-	-	.0012	5.5	-	3
*†					.010	4.5	-	4
*†.05	.90	.05	-	-	.0050	3.5	-	5
*†					.038	2.5	-	6

A star indicates that results are suitable for use of the variability chart alone, the c + a chart.

A dagger indicates that the results are suitable when both the mean and variability charts are used together.

7. Miscellaneous comments and conclusions. If the frequency function is not normal, a transformation should be used to make the frequency function approximately normal. Even in cases of very skewed distributions or of pathological distributions the (c - a, c + a) charts will work where the $(\bar{x},R)$ charts might fail. The O.C. curves for the (c + a, c - a) charts would be found just as in this table, but a comparison against the "best" O.C. curves would depend on a knowledge of the distribution of the particular statistics.

If larger sample sizes are required two options are open. Calculate the O.C. curves by use of the exact distribution of c - a, or assume that the distribution of c - a is approximately normal with mean $n(r - p)$ and variance $n\{(p + r) - (p - r)^2\}$. Sample sizes of n = 16 and n = 20 might prove useful. The choices we have made are certainly adequate for most needs and are much better than the usual small sample sizes of 5 used for the $(\bar{x},R)$ chart. It should be clear that gauging will not entail measuring each item twice, since the proportion exceeding x_2 is r and the proportion less than x_1 is p. So far as costs are concerned gauging will be cheaper, because of the ease of gauging and because counting is cheaper and less subject to error as compared with the arithmetic involved in the $(\bar{x},R)$ charts. We do not advise the use of gauging when the process has not been brought into control. In such cases the $(\bar{x},R)$ charts are preferable. If we know the proportion of items below x_1, and the proportion of items above x_2, and if we assume that the items are distributed normally we may estimate $\bar{x}'$ and σ'.

by

$$\sigma = \frac{x_2 - x_1}{u_2 - u_1}$$

and $$\bar{x} = \frac{x_1 + x_2}{2} - \frac{(u_1 + u_2)}{2}\,\sigma,$$

where u_1 is determined by

$$\int_{-\infty}^{u_1} \varphi(t)\,dt = \frac{a}{n},$$

and u_2 is determined by

$$\int_{u_2}^{\infty} \varphi(t)\,dt = \frac{c}{n}, \quad \varphi(t) = \frac{e^{-t^2/2}}{\sqrt{2\pi}},$$

$$n = a + b + c.$$

References

(1) Stevens, W. L., "Control by Gauging," Journal of the Royal Statistical Society, Series B, Vol. X, pps. 54-108, 1948.

(2) Computation, Laboratory of Harvard University "Tables of the Cumulative Binomial Probability Distribution," Harvard University Press, Cambridge, Massachusetts, 1955.

(3) Pearson, E. S. and Hartley, H. O., Editors, "Biometrika Tables for Statisticians, Vol. I," Cambridge University Press, London, England, 1954.

(4) Ott, Ellis R. and Mundel, August B., "Narrow-Limit Gauging," Industrial Quality Control, Vol. X, No. 5, March 1954.

ARMY PROGRESS IN QUALITY CONTROL DURING CALENDAR YEAR 1958

Colonel E. A. Koerner, Chief, Standards Br., Procurement Div.
Office of the Deputy Chief of Staff for Logistics

Mr. Chairman, ladies and gentlemen.

When the Department of the Army was asked to present a paper at this convention, the title of the paper specified was "Army Progress in Quality Control During Calendar Year 1958". As a springboard for a discussion of progress in quality control during 1958, let's review quality control as a concept.

A most important consideration in such a review regards a realization that ever since man has made things, even as far back as the first club or stone axe, he has had to exercise quality control. This quality control depended upon standards of measurement and became increasingly important as we developed more complex items with higher requirements for interchangeability. One fact, not too well known to many people, is that Eli Whitney, who first proposed the concept of interchangeable parts to allow mass production to be used could never successfully implement his concept. The first actual use of the concept of interchangeable mass production was probably made by John Hall in his rifle plant at Harper's Ferry in the year of 1819. It was possible for him to do so because he had precise enough standards of measurement to allow an adequate quality control system to be established. True, this quality control system did not use statistical techniques, did not depend on records and highly integrated organizations for feedback, but it was quality control nevertheless. The standards of measurement which enabled accurate gages to be produced allowed control of machine production quality through the first piece checks, and control of overall output through a screening inspection of the parts produced.

From this early start of a quality control system, mass production systems have been subject to quality control of a type which did not know the niceties of statistical techniques but did depend upon people and craftsmen, who could manage quality in an adequate manner. This condition existed in quality control up to and into the early days of World War II. A serious shortage of craftsmen and a serious shortage of gage making capacity made it necessary to control quality with something other than 100% gaging of parts.

The tool of statistical quality control that had been proposed many years earlier by Walter Shewhart and that had laid dormant for a couple of decades, except for Bell Laboratories and Western Electric useage, was dusted off, sharpened up and put to work by Army Ordnance. This tool has been developed into a highly discriminating series of statistical techniques for evaluating quality and making decisions as to its adequacy.

Army, Navy, and Air Force and many US Industries, as well as industry in many other countries, make use of this technique today. I believe that the availability of this tool coupled with the idea of the military inspector measuring quality himself led us into

faulty organization. We all know that quality of an item should be controlled during its production and not afterwards. It hasn't always been clear that such control is a part of the production program and for that reason should be exercised by contractors and not by Government inspectors. The Department of the Army and other military departments for the past few years have been firming up policy in this area. We still have the old procedure of Army inspection for use when there is no other way to get adequate quality. Even when this is accomplished by statistical sampling it is, in essence, a screening of quality after the fact of production.

For this reason, the Army has established a policy stressing two of four different types of inspection. The first two called types A and B in our Army Regulations 715-20 are the old style Army inspection procedures. The last two, types C and D stress the contractors responsibility for controlling production quality during production. They place the burden of operating details, and authority for control of these operations, with the contractors.

You may ask at this point, "What is the purpose of the Government quality control man"? This can be simply stated. His role is that of determining if the contractor has operated an adequate quality control program and, on the basis of such determination to make the decision to accept the product offered by the contractor to the Government.

The difference between the two types of inspection, types C and D mentioned earlier has to do with the type of items and type of contracts which vary due to differences in the items. When we can clearly describe the required items in specifications and can write adequate quality assurance provisions based on inspection of completed components or items, we will require that contractors perform as a minimum these quality assurance provisions or their equivalent. Type C inspection goes beyond this with the intent of persuasion of contractors to develop control systems equivalent to these quality assurance provisions in terms of control of quality during production. When this happens, the number of people Army must have on its payroll can be significantly reduced. Type D inspection is used on contracts for complex items which cannot be adequately described in detail specifications and for which adequate quality assurance provisions cannot be developed. This type D inspection depends upon a quality control system made mandatory by contract in accordance with a DOD quality control system requirements specification.

As I told you to begin with, I was asked to come here to tell you about Army progress in quality control during 1958. Thinking first of progress in terms of use of types C and D inspection, I can report some significant advances. Contractors in 650 plants have successfully established quality control and inspection programs which allow Army to use type C inspection. Type D inspection programs have been established in about 75 plants. The balance sheet at the end of the year shows that we are handling a workload slightly higher than that of about a year ago with 250 less people. A conservative estimate of the salary savings alone is one million dollars or more. Savings due to reduction in loss of production manhours and in materials cannot be estimated. They probably outweigh the salary savings.

These savings that I report are small in relation to what we anticipate is the potential of the program. The potential is probably better illustrated by the success of one Army Technical Service which was able to handle a 40% increase in workload with personnel reduced by 5%.

The progress I have reported so far has been on that part of quality control which has only had to do with industrial production operations and determinations that products were satisfactory. Actually quality control that you in Industry are interested in starts much earlier than your production process. It starts at the time of deciding upon military characteristics and remains with the programs continually through research and development, proto-type production, preparation of specifications and even has an impact on the way we write our contracts. Our gains in the other facets of quality control are not as outstanding as the gains made by the Industry-Army team covering production. In order to assure the preparation of more adequate specifications, we are still running a military specification course at the Army Logistics Management Center, Fort Lee, Va. In order to up-grade our personnel in the use of types C and D inspection, we are developing a training course on quality control management. This training course will orient people in ways of assessing organization for production, organization for quality control, and the relations between these two, and methods of evaluation of production quality control programs in a general sense.

During the coming year, we are going to emphasize the expansion of types C and D inspection in many more plants than the 650 of type C and 75 type D that was the status as of the end of last year. We are going to continue to press for better and better specifications. We are going to expand this push to cover other documents having an impact on quality control. Some of these documents you see in the procurement contracting system. These are specifications, standards, and in some instances handbooks. Others, which have to do with technical committee actions and the establishment of military characteristics, never do get into the procurement data package cited in your contracts. Nevertheless they have a vital impact on this quality control program and are thus in need of review to see that they mesh properly in the overall Industry-Army quality control system.

One more point which has high importance in quality control has to do with feedback. Feedback of information to allow correction of a system is used almost intuitively in many cases without realizing that it is feedback. The simplest example being the setup man who goes by a machine once in a while, checks the piece and adjusts the machine. A more complex version of the feedback that is not resulting in corrective action as it should, has to do with the information that you in industry have regarding the ways of making the things that we ask for in our specifications, and the faults in our specifications. For years we have been trying to tap this fund of information in order to develop better specifications. We haven't had as much success as we would like and as is possible. Tremendous strides in improving the effectiveness of our quality control program can be made if you freely feedback the information you have at the time we coordinate specifications with industry.

Let us sum up where we have been, where we are, and where we are going. You in industry have always had quality control. In many cases a decade or so ago your controls, even though adequate, were not of a sort that could be used directly by the Army in making acceptance decisions. We have learned together how to correct this situation. Today, under Army policies most industries voluntarily develop quality control programs that fit their own situations. Usually the only adjustments in these programs to make them suitable for Army are clarifications of intent both by Army and by industry. As we gain experience the need for adjustments will gradually disappear. The key point I want to leave with you is Army's strong conviction that voluntary industry development of quality control programs tailored to fit each different situation will give maximum efficiency in our procurement.

I'm sure that this year and in future years we will expand the use of industry's quality control to the benefit of all of us. These benefits will total many times the million dollars I mentioned before.

PROCESS CONTROL OF CONTINUOUSLY EXTENDED PRODUCT

Dorsey E. Dean
The Martin Company

The outstanding characteristic of continuously extended product is its lack of discrete units of production. In this type of production, for Quality Control purposes, a production unit is much less clearly defined than in the usual type consisting of distinct pieces, parts, components, or assemblies produced on a repetitive basis.

The lack of a clearly defined production unit presents no problem in the manufacturing and production control departments since in these areas, units of continuously extended product can be arbitrarily and conveniently defined as discrete lengths, pieces, or rolls. For Quality Control purposes, however, such distinctions are artificial and may or may not be related to the pattern of process variation.

Therefore, to be effective on continuously extended product, Quality Control must monitor and control, within economically feasible limits, the continuous variation within the process itself rather than that occurring from arbitrary unit to arbitrary unit. In the usual discrete type of production on which most of our Quality Control theory is based, this is covered automatically. Through use of small sample techniques based upon consecutive units of production, Bar X and R Charts indicate variation within each sample as well as that occurring from sample to sample.

If we are to effectively apply previously developed small sample techniques, the basic question which must be resolved is, "What constitutes a representative sample of continuously extended product?" In the author's opinion, the answer to this question is a direct indication of the effectiveness of the Quality Control program on any continuously extended product.

The simplest and most erroneous answer is to use the arbitrary units established by the manufacturing departments using Bar X and R Charts to show variation in sample averages and from unit to unit in each sample. The error in this approach is the fact that variation within a single unit of the sample is completely ignored. The Bar X using this approach is more analogous to a double Bar X than it is to a true Bar X and the range is correspondingly more analogous to an average range than it is to a true range.

This approach, in effect, averages most of the variation out of the process and as a result, produces a false sense of security. An average without a valid measure of dispersion is for Quality Control purposes, practically worthless. Misapplication of statistical quality control techniques in this manner can kill a Quality Control program just as the non-swimmer was drowned hurriedly wading across a river having an average depth of two feet. The one area ten feet deep killed him.

The problem then becomes one of measuring and controlling dispersion within the continuously extended product. Since use of an arbitrary production unit ignores dispersion within the unit itself, it is obvious that the samples must be taken within the unit if a true picture is to be

obtained and the source of variation determined and controlled.

Depending upon the quality characteristic under consideration, a given sampling frequency may or may not be satisfactory. Since most continuously extended product processes tend to be cyclic in nature, any sampling plan must be designed with this in mind. The sampling plan should evaluate both amplitude and frequency of the cycle if control is to be achieved.

At the opposite extreme of taking samples in such a manner as to average the variation out of the process is taking small single samples within the arbitrary production units and making manual adjustments to the process on the basis of each sample. In a cyclic process, such a procedure, in most instances, will increase rather than reduce variation. Most single sampling of this nature is done by production operators who in trying to do a conscientuous job tend to overcontrol the process through continuous adjustment of settings.

If the operator is aware of the cyclic feature, he might not compensate for it properly. If he is not aware of the cycle and the sample is taken at an extreme high or low point, the result is a continuous shift in the midpoint of the cycle. If the cycle midpoint is at the specified value and its extreme points are just inside the upper and lower tolerances, adjustment of the process on the basis of this single misleading sample will produce out of tolerance material.

On the other hand, if the cycle is averaging higher or lower than the specified value, the amount of adjustment will only be correct if the sample is taken at this midpoint in the cycle. Since it is highly improbable that the sample will be taken at this point, depending upon the point sampled in the cycle, the correction will be too small, too great, or conceivably in the wrong direction. Obviously, use of a single sampling method on a continuously extended product process will result in an increase in variation due to "hunting" except under extraordinary circumstances.

From this discussion, it is obvious that if statistical quality control techniques are to be effectively applied to the production of a continuously extended product, a sampling plan must be devised which will simultaneously prevent overcontrol and averaging the variation out of the process. In order to satisfy both requirements, sampling should be as frequent as possible. This leads to the logical conclusion that a continuously extended product should be continuously measured if optimum control is to be achieved.

The continuous measurement should be made on an automatic basis through the use of mechanical, pneumatic, photoelectric, dielectric, magnetic or radioactive measuring instruments rather than by manual means and should be coupled with an automatic recorder. This will permit immediate detection of any cycles or trends developing in the process and provides a sound criterion for manual adjustment of the process.

On processes where the adjustment of the measured variable is limited as is the case in the continuous extrusion of a given contour through a fixed die, continuous measuring and recording of process variables is

still desirable as a means of indicating when reworking of the die is necessary. On this type of application, savings realized through reduction in unwarranted reworking of dies can often more than justify the cost of the automatic measuring and recording instruments.

The conclusion that continuously extended product should be continuously measured for optimum control would appear to be a statement from the old school which advocated 100% inspection rather than from the new school of statistical quality control. However, this concept differs from the old school philosophy in that the measuring and recording are:

1. Automatic rather than manual,
   and are on a
2. Variable rather than attribute basis.

A further difference is that the continuous chart record itself can be used as a basis for application of statistical quality control methods. Intensive analysis of continuous chart records on a statistical basis is an ideal method for evaluating machine capability for a continuously extended product while statistical sampling of these charts is a very convenient method for evaluating the workmanship of different operators.

A practical extension of the automatic continuous measuring and recording concept would be to couple the continuous measurement with an automatically controlled means of adjusting the measured parameter to its specified value or setpoint. For example, if a material is being continuously rolled to a specified thickness and a measuring device senses that the thickness is greater than specified, control action is automatically initiated to decrease the setting of the rolls.

Since, as previously discussed, most continuously extended product processes tend to be cyclic in nature, this factor must be taken into consideration when coupling a continuous measurement with an automatically controlled means of adjusting the measured parameter. Unless this is done, an automatic adjusting mechanism can increase rather than reduce variation through excessive "hunting". The effect is similar to that previously discussed in connection with manual adjustment on the basis of single samples. "Hunting" of the automatic adjustment can be eliminated by damping the adjustment or by basing the adjustment upon a moving average rather than an instantaneous value.

In our example of rolling a continuous length to a specified thickness, cycling of the thickness could be caused by non-uniform roll surfaces or an elliptical circumferential contour produced in the roll during an extensive shutdown. Since under these circumstances the material thickness will vary with each rotation of the roll, a moving average will eliminate the cycle for adjustment purposes and still correct any trends in the thickness setting.

In summary, the lack of a discrete production unit and a cyclic tendency have been discussed as two characteristics of continuously extended product. It has been indicated that these characteristics preclude the indiscriminate application of statistical quality control methods. Automatic continuous measuring and recording have been advocated as

the most desirable means of indicating process variation. Sampling of recorder charts has been recommended as a method for applying statistical quality control methods to continuously extended product while coupling of the continuous measuring and recording to automatic adjustment of the measured parameter has been discussed as the ideal method of control.

SUMMARY OF EXTREME-VALUE THEORY AND ITS RELATION TO RELIABILITY ANALYSIS

James R. King
Senior Engineer
Research and Advanced Development Division
Avco Manufacturing Corporation

I. SUMMARY OF EXTREME-VALUE THEORY

A. Historical Background

The known part of the question of how to deal with values which lie well away from the central value of a set of data goes back at least to the time of Gauss. The concern over "outliers" in astronomical data lead to a solution in the "theory of errors." Statistical methodology eventually adopted the Normal Error Curve. Use of the normal distribution parameters gives various error ratios, confidence intervals, and probability levels to be used in testing "unusual values."

More recently, in the 1920's, as a consequence of the preoccupation of statisticians with sampling distributions, the estimation of distribution parameters and their estimation errors received great attention. Included in this work were studies, from a theoretical standpoint, in the sampling distribution of the range and of the largest normal value of the normal distribution, which indicated that the range and largest value are variates in their own right with their own distributions. It was further determined that the parameters of these distributions vary with sample size.

In the 30's, practical applications for extreme value statistics were found in human life statistics, radioactive emission, and strength of materials. In the 40's, further applications were made to floods, meteorology, capacitor design, fatigue of materials, and aircraft gust loads.

Theoretical development has continued and, in the 50's, increasing interest has led to verification of key extreme value distribution characteristics through advances in rank-order statistics and maximum liklihood estimation technique. These factors coupled with information available from the mathematical analysis of exponential-type distributions have extended the range of applications. Useful applications have been in life-testing, reliability estimation, business operations, and transportation maintenance systems.

B. Modern Development

The developments which led to a particular theory of extreme values were due in large part to Dr. E. J. Gumbel and are summarized in his NBS lectures (1). This has been augmented by Lieblein (2). A more complete development has been recently published by Gumbel (3) which includes an exhaustive bibliography.

The title of "extreme-value theory" is derived from the fact that the necessary statistical theory was first developed for answering questions about the sampling distribution of extreme values. However,

the statistical work has led to an understanding of extreme values in the light of exponential type distribution characteristics and some properties of order statistics. This new knowledge of exponential distribution and order statistics extends the range of data analysis by offering new distributions and characteristic properties from which to choose in forming hypotheses. In this way, therefore, the title of this branch of statistics tends to mask its wide domain in the area of modern statistical analysis.

C. A Method of Application

A convenient method of analyzing any distribution is to plot it on an appropriate probability paper, to determine a best fitting regression line with confidence limits, and by comparison of the plotted points to the fitted line and confidence limits to judge the appropriateness of the hypothesis of distribution assumption made in choosing the probability paper. Thus, a normal distribution plots as a straight line on arithmetic probability paper and a log-normal distribution is straight on log-normal probability paper. Gumbel has devised a probability paper on which extreme value distributions plot as straight lines. In testing a new or unknown distribution by graphical means, the use of Gumbel's paper can extend the range of investigation.

On the other hand, since extreme value distributions are skewed, one might first test a skewed distribution on Gumbel's paper. Some of the typical distributions considered as extreme value types by Gumbel are chi-square with a small number of degrees of freedom, log-normal, logistic, range and the exponential as the arch-type. The normal distribution is also included but behaves rather oddly. In particular, its extreme values converge to their asymptotic values very slowly. An example using extreme-value probability paper is given in the Appendix.

D. Discussion

A statistical theory of extreme values is desirable to explain extreme values observed under a given set of conditions and to predict the extremes which may be expected under the same or an equivalent set of conditions. The following requirements are necessary:

1. a statistical variate,

2. that the underlying distribution does not change from sample to sample, and

3. that the data are independent.

Two possible sources of data exist. First, a set of data may be obtained by extracting one extreme value from many equivalent collections of sample data. For example, during one year a river will have 365 maximum-flow rates at a given station. The largest of all these is designated as the annual flood. The annual flood from a number of years is treated by extreme-value theory, and in fact, has led to some of the earliest practical applications by Gumbel (1). Further examples are given in his paper of applications to meteorological and geophysical data, breaking strengths, aeronautical gust loads, and life testing.

Second, for certain distributions, which are generalized by Gumbel as Type I, of the form $\exp(-e^{-y})$, it is possible to apply extreme-value theory to a set of data which does not necessarily consist of selected extremes. Botts (4) illustrates this kind of application to annual loss costs of a mutual crop-hail insurance company.

Gumbel's Type I includes the normal, exponential, chi-square, log-normal, logistic, and range distributions. These represent a large percentage of distributions achieved in physical testing and to many of the failure patterns attributable to environmental testing.

Of particular interest is the fact that these distributions, with the exception of the normal, are generally highly skewed. As a matter of practical consideration, most distributions found in testing tend to be slightly to moderately skewed.

Gumbel also identifies two other types of equations. Type II is the Cauchy type of the form $\exp[-(v_n/z)^k]$. Type III is of the form $\exp\left[-\left(\frac{\omega - x}{\omega - v_n}\right)^k\right]$ and is frequently called the Wiebull distribution.

There are some other considerations which may suggest the need and the areas for using extreme-value theory.

First, products which are manufactured under a regime of effective quality control may tend to become skewed as control criteria become better known and are better applied.

Second, well-designed precision products may tend to yield skewed distributions of data when in a technical state of control.

The two preceding considerations can result because the freedom of a product to vary is increasingly constrained by more critical knowledge of the criteria for maintaining compliance to specification requirements. When a product has considerable latitude to vary, random influences and averaging effects tend to combine and to approach normal distribution characteristics. This has long been explained by the central-limit theorem, (5) which states that the sum of effects of many independent random variables approaches a normal distribution as the number of variables increases. This approximation is even quite good when the number of variables is only four nearly equal independent variables.

The individual sources of variation are frequently considered as individual variables. However, in complex products, these sources may actually be sets of dependent variables which may reduce the number of independent factors considerably. In addition, as statistical control criteria are applied some of these sets may be identified as assignable causes of variation, investigated, and substantially reduced or eliminated. This further reduces the number of independent sources. Thus, the number of independent effects may eventually become as small as 2 or 3. At this point, it is possible that only one of the variables remaining is of sufficient influence to affect the product significantly. Now, an accurate description of this variable may well be close to an analytic description of the product or process function itself and this is not necessarily normal.

Alternatively, if the elimination of independent assignable causes leaves a set of nearly equal but dependent variables, the result is to approach a log-normal distribution.

Good precision product design accomplishes effectively the same things by minimizing the influence of all but a very few fundamental parameters.

As a third consideration, many electronic and electromechanical devices are inherently function generators of a generally exponential type. It is possible in many cases that random variations in manufacturing tolerances, methods, fits, alignment, etc., may be transformed by the device to an exponential distribution of observed output error. Logarithmically wound potentiometers are an example. Ferrell (7), (8) illustrates some results for transistor currents using log normal paper and makes several interesting points with regard to choices of hypotheses and the variation in conclusions drawn as a result of different hypotheses.

Last, with regard to failure patterns, Davis (6) states that: ". . Those systems which exhibit reasonable agreement with this (exponential) failure theory are characterized by a predominance of human errors as the cause or a careful and well-developed operating technique for minimizing failure. Systems which are subject to a wide range of environmental severity also appear to follow this pattern.

"The normal theory of failure appears to apply to systems which exhibit small variations in failure resistance among the individuals within a population and which are subject to small variations in environmental severity. Further, the failure resistance of the mechanism deteriorates with time and operational procedure requires that each item be used until ultimate failure. . ."

II. APPLICATIONS TO RELIABILITY ANALYSIS

A. Definitions

For the purpose of our discussion, we must start out with a working definition of reliability analysis. Let us base this definition on the two most distinct areas of reliability analysis:

1. Engineering reliability analysis is the usage of acquired data to evaluate design, manufacture, and test of items or systems with respect to a pre-determined or desirable standard of performance.

2. Statistical reliability analysis is the usage of data to establish models of behavior, to test various possible hypotheses and to estimate error magnitudes using both statistical and mathematical techniques for the primary purpose of predicting the probabilities of successful performance of the items or systems.

Both kinds of reliability analysis could be performed, and in many cases, are being performed, by the same person or group of persons. Since the engineering questions permit statistical answers and the formulation of statistical hypotheses requires engineering judgements, there is a lot of overlap which can easily generate confusion.

We must then remember that the engineer wants to evaluate actual performance with respect to some standard for the purpose of determining the validity of the engineering concepts involved. The statistician, on the other hand, has the problem of determining objective probabilistic measures of achieving or continuing to achieve some level of performance with respect to an assigned standard.

B. Data

Since we have based the definitions of reliability analysis in part on "usage of data;" let us consider the kinds of data available.

1. Discrete data

   a. Relatively inexpensive, easily repeated.

   b. Relatively expensive, not readily repeated.

2. Continuous data

   a. Inexpensive, repeatable

   b. Expensive, non-repeatable

The analysis connected with relatively inexpensive easily repeatable data poses no special burden. On the other hand, relatively expensive data which is not readily repeated requires analysis commensurate with its cost and importance. This may require forms of analysis not now generally in use. Extreme-value theory is one of these forms and when its hypotheses can be shown to be more valid than more common forms, it should be used.

In many reliability analysis situations, the measurement of central tendency may not be particularly important because we may be dealing with such questions as: How often will a certain parameter exceed a given value? What is the chance of an environmental stress being as great as the design limit for a device or system in a specific application?

Continuous data records are frequently collected for use in such determinations and the analysis of such records can be simplified by analyzing extremes in appropriate time intervals. This may improve accuracy of estimation in cases where the data processing may pose problems in filter selection and bandwidth, frequency response, and signal-to-noise ratio. This is frequently a problem for values of small magnitude in a distribution.

In life testing, Epstein (9) uses the exponential model to derive life test distribution models which include the extreme value and the Weibull distributions. Kao (10) has used the Weibull distribution in analyzing electron tube life tests.

In summary, we have considered the areas of use of extreme-value theory in:

1. Design analysis and evaluation by engineering reliability testing and analysis.

2. Reliability prediction as a function of statistical reliability analysis.

3. Prediction of failure rates or levels by statistical analysis with consideration of data reduction requirements.

4. Testing distributions for exponential behavior.

5. Better fitting of skewed distributions.

6. Judging "unusual" values as a function of extending the range of choice of hypotheses for consideration.

REFERENCES:

1. Gumbel, E. J., Statistical Theory of Extreme Values and some Practical Applications, NBS Appli. Math. Ser. 33, 1954.

2. Lieblein, J., A New Method of Analyzing Extreme-Value Data, NACA, TN 3053, 1954.

3. Gumbel, E. J., Statistics of Extremes, Columbia University Press, 1958.

4. Botts, R. R., Extreme-Value Methods Simplified, Agr. Econ. Research 9 (3), 88-95, July, 1957.

5. Cramer, H., Mathematical Methods of Statistics, Princeton University Press, Princeton, pp. 213-221, 1951.

6. Davis, D. J., An Analysis of Some Failure Data, J. Am. Stat. Assoc. 47, 113-150, 1952.

7. Ferrell, E. B., Normal or Log-Normal Probability Paper for Plotting Experimental Data, Industrial Quality Control, Vol. XV, No. 1, 12-15, 1958.

8. Ferrell, E. B., Log-Normal Control Charts, IQC, Vol. XV, No. 2, 4-6, 1958.

9. Epstein, B., The Exponential Distribution and Its Role in Life Testing, Industrial Quality Control, Vol. XV, No. 6, 1958.

10. Kao, J. H. K., The Weibull Distribution in Life Testing of Electron Tubes, presented at the 115th Annual Meeting of The ASA, 1955.

APPENDIX

USE OF EXTREME-VALUE THEORY AND DOUBLE EXPONENTIAL PROBABILITY PAPER

The following procedure follows that given by Botts (4). Some tables have been added to simplify the task further for users.

PRELIMINARY

When a set of data has been collected, construct a frequency diagram to determine whether the data are skewed toward the larger or smaller values of the measured variables. If skewed toward larger values, prepare a table ranking the data in order from smallest to largest; if skewed toward the smaller values, rank the data from largest to smallest. Compute the arithmetic mean and the standard deviation of the set of data. This is shown in Table I, column (2), for data on the current which is required to burn out the fusible link in a thermal relay. Also shown is the rank m for each value, column (1), and the plotting position, column (3).

The ranks and, consequently, the plotting positions vary with sample size. Table II contains plotting positions for a number of sample sizes. Any other set of values which is needed may be calculated readily from $\frac{m}{n+1}$ for a specific n.

REGRESSION LINE

The equation for the regression line is determined in three parts.

First, calculate the slope, (4) in Table I, which is the standard deviation of the sample divided by the expected standard deviation for sample size n, σ_n (Table III). Next, (5) in Table I, the slope, $\frac{1}{\alpha}$, is multiplied by the expected mean for sample size n, $\bar{y}_n$, also from Table III. Last, subtract this product, (5), from the arithmetic mean to determine the mode, μ, (6) in Table I.

The equation of regression, (7) in Table I, is then $X = \mu + \frac{1}{\alpha}\bar{y}_n$ where μ = mode of the density function and $\frac{1}{\alpha}$ is the slope. $\bar{y}_n$ is the expected mean of the reduced variate, $y = \alpha(x-\mu)$, where x is the original variate and μ is the expected largest (or smallest) value.

The plotting position and the reduced variate, y, are related by $\frac{m}{n+1} = \exp(-e^{-y})$. The derivative of $\exp(-e^{-y})$ has a maximum at $y = 0$ which precisely defines the mode. The slope $\frac{1}{\alpha} = \frac{s_x}{\sigma_n}$, is determined directly from the sample standard deviation, s_x, and the expected standard deviation, σ_n.

Three plotting points, (8) in Table I, are determined for the regression line. Although two would be sufficient, three serve as a convenient check on the calculations. The same probability values can be

used repeatedly and appear in Table IV. The probability values chosen are 0.20, 0.50, and 0.95. The regression line may now be drawn on extreme-value probability paper (Fig. 1). The observed variable is plotted vertically, and the cumulative probability horizontally. The original data may now be plotted against the tabulated plotting positions.

TESTING FIT OF DATA

The fit of the data to the line may be established by use of control lines which are computed as follows:

Divide the slope, $\frac{1}{\alpha}$, by the square root of the sample size, $\sqrt{n}$, as in (9) and multiply by the factors given in Table V for $\frac{m}{n+1}$ when this value lies between 0.15 to 0.85.

For the second largest and largest values obtained, multiply the slope directly by the last two factors shown in Table V. The control values so obtained are set off as vertical distances above and below the regression line at the indicated probability points up to 0.85 and at the plotting positions for the two largest values. Again, the same multipliers may be used repeatedly. The control lines must enclose two-thirds of the points for extreme-value theory to be applicable. If not, reject the assumption of applicability.

The foregoing values for $0.15 \leq P \leq 0.85$ derive from a theorem by Laplace (1) which may be formulated as

$$\sqrt{n}\ \sigma(y_m) = \frac{\sqrt{\Phi(y)\,[1 - \Phi(y)]}}{\phi(y)}$$

where $\Phi(y)$ is the cumulative probability function,

and $\phi(y) = \Phi'(y)$ is the probability density function, and

$\sqrt{n}\ \sigma(y_m)$ is a pure number which is the standard error of the reduced variate, y.

This permits us to calculate $\sigma(x_m) = \frac{\sqrt{n}\ \sigma(y_m)}{\sqrt{n}\ \alpha}$

Gumbel derives the control interval for the two largest values as

$\Delta\ x_{n-k} = \frac{y(P_{0.683})}{\alpha_{n-k}}$, where k is 0 or 1 for the largest and second largest values respectively; $y(P_{0.683})$ is the value of the reduced variate y for which $P = 0.683$, given as 1.14071 and 0.75409, respectively; and $\alpha_{n-k} = \alpha$.

The 68 percent confidence interval for a probability estimate from an extreme-value plot is 0.32T to 3.12T, where $T = \frac{1}{1 - P}$.

In cases where it may be of interest to compare failure rates at the same level of a variate, x, between two unequal sample sizes from the same population or to estimate expected failure rates for future sample sizes from a known sample, the following relationship is helpful:

$$x_1 = \mu_1 + \frac{\ln T_1}{\alpha_1}, \quad x_2 = \mu_2 + \frac{\ln T_2}{\alpha_2}$$

where ln is the natural logarithm, $T = \frac{1}{1 - P}$,

$$\mu = \text{mode}, \quad \frac{1}{\alpha} = \text{slope},$$

$$\text{for } x_1 = x_2, \text{ then, } \ln T_2 = \alpha_2 (\mu_1 - \mu_2) + \frac{\alpha_2}{\alpha_1} \ln T_1$$

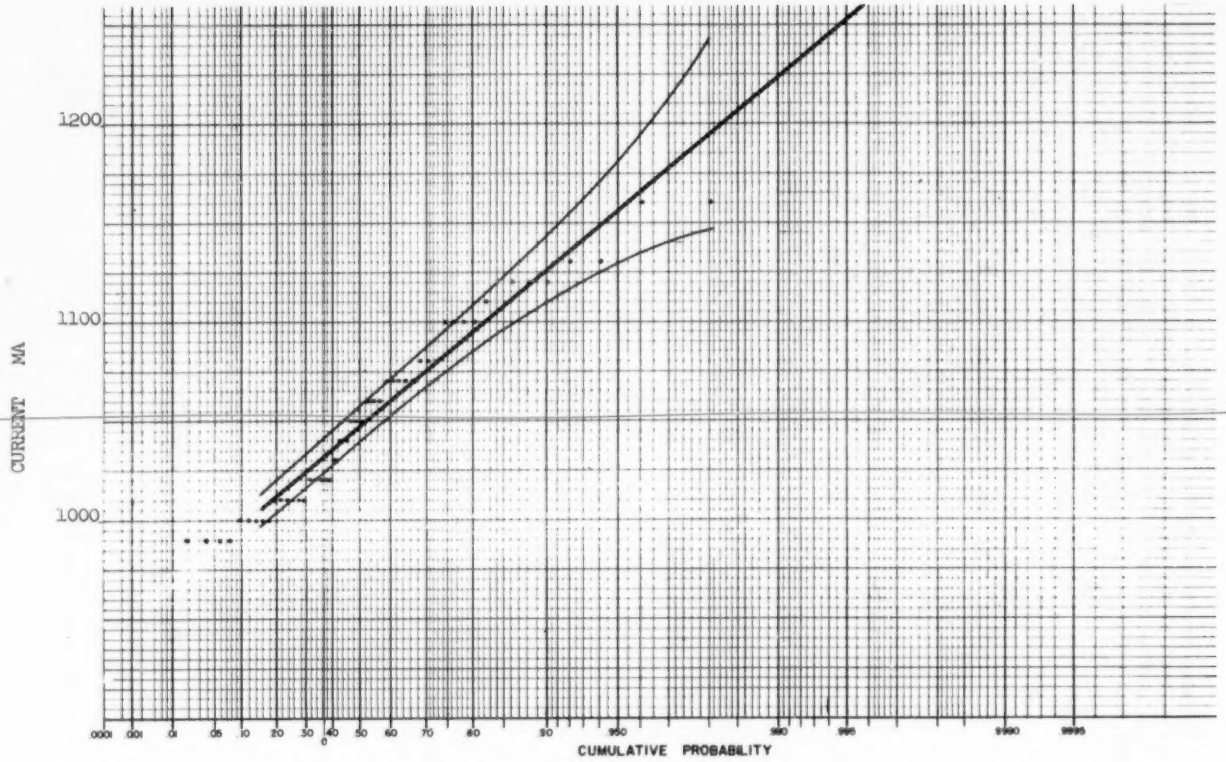


Figure 1 CURRENT REQUIRED TO BURN OUT A FUSIBLE LINK

TABLE I

SAMPLE CALCULATION SHEET

(1) Rank m	(2) Current ma	(3) Plotting Position $\frac{m}{n+1}$	(1) Rank m	(2) Current ma	(3) Plotting Position $\frac{m}{n+1}$
1	990	0.0196	27	1060	0.5294
2	990	0.0392	28	1060	0.5490
3	990	0.0588	29	1060	0.5686
4	990	0.0784	30	1070	0.5882
5	1000	0.0980	31	1070	0.6078
6	1000	0.1176	32	1070	0.6274
7	1000	0.1372	33	1070	0.6471
8	1000	0.1569	34	1070	0.6667
9	1000	0.1765	35	1080	0.6863
10	1010	0.1961	36	1080	0.7059
11	1010	0.2157	37	1080	0.7255
12	1010	0.2353	38	1100	0.7451
13	1010	0.2549	39	1100	0.7647
14	1010	0.2745	40	1100	0.7843
15	1010	0.2941	41	1100	0.8039
16	1020	0.3137	42	1110	0.8235
17	1020	0.3333	43	1120	0.8431
18	1020	0.3529	44	1120	0.8627
19	1020	0.3725	45	1120	0.8824
20	1020	0.3922	46	1120	0.9020
21	1030	0.4118	47	1130	0.9216
22	1040	0.4314	48	1130	0.9412
23	1040	0.4510	49	1160	0.9608
24	1050	0.4706	50	1160	0.9804
25	1050	0.4902		Mean 1054.4	
26	1050	0.5098		Standard Deviation 48.62	

TABLE I

SAMPLE CALCULATION SHEET (cont'd.)

(4) Slope = $\frac{48.62}{1.1607} = 41.88$

(5) $(41.88)\ (0.5485) = 23.0$

(6) Mode = 1054.4 - 23.0

= 1031.4

(7) Regression Equation

X = 1031.4 + 41.88 y

(8) Plotting Points

$\frac{m}{n+1}$	y	(slope) x (y)	x
0.20	-0.47588	- 19.9	1011.5
0.50	+0.36651	+ 15.4	1046.8
0.95	+2.97020	+ 124.4	1155.8

(9) $\frac{(4)}{\sqrt{n}} = \frac{41.88}{\sqrt{50}} = 5.92$

$\frac{m}{n+1}$	Factor	x (9) or (4)	Control Interval
0.15	1.255	x 5.92	7.4
0.50	1.443	x 5.92	8.5
0.70	1.835	x 5.92	10.9
0.85	2.585	x 5.92	15.3
2nd	0.75409	x 41.88	31.6
1st	1.14071	x 41.88	47.8

TABLE II

PLOTTING POSITIONS

n = 15	16	17	18	19	20	25	30
0.0625	0.0588	0.0555	0.0526	0.0500	0.0476	0.0385	0.0322
0.1250	0.1176	0.1111	0.1053	0.1000	0.0952	0.0769	0.0645
0.1875	0.1765	0.1667	0.1579	0.1500	0.1429	0.1154	0.0968
0.2500	0.2353	0.2222	0.2105	0.2000	0.1908	0.1538	0.1290
0.3125	0.2941	0.2778	0.2632	0.2500	0.2381	0.1923	0.1613
0.3750	0.3529	0.3333	0.3158	0.3000	0.2857	0.2308	0.1935
0.4375	0.4118	0.3889	0.3684	0.3500	0.3333	0.2692	0.2258
0.5000	0.4706	0.4444	0.4210	0.4000	0.3810	0.3077	0.2581
0.5625	0.5294	0.5000	0.4737	0.4500	0.4286	0.3462	0.2903
0.6250	0.5882	0.5555	0.5263	0.5000	0.4762	0.3846	0.3226
0.6875	0.6471	0.6111	0.5789	0.5500	0.5238	0.4231	0.3548
0.7500	0.7059	0.6667	0.6316	0.6000	0.5714	0.4615	0.3871
0.8125	0.7647	0.7222	0.6842	0.6500	0.6190	0.5000	0.4194
0.8750	0.8235	0.7778	0.7368	0.7000	0.6667	0.5385	0.4516
0.9375	0.8824	0.8333	0.7895	0.7500	0.7143	0.5769	0.4839
	0.9412	0.8889	0.8421	0.8000	0.7619	0.6154	0.5161
		0.9444	0.8947	0.8500	0.8095	0.6538	0.5484
			0.9474	0.9000	0.8571	0.6923	0.5806
				0.9500	0.9048	0.7308	0.6129
					0.9524	0.7692	0.6452
						0.8077	0.6774
						0.8462	0.7097
						0.8846	0.7419
						0.9231	0.7742
						0.9615	0.8064
							0.8387
							0.8710
							0.9032
							0.9355
							0.9677

TABLE III

EXPECTED MEANS AND STANDARD DEVIATIONS
OR REDUCED (y) VALUES BY SAMPLE SIZE

Sample Size n	Average $\bar{y}_n$	Standard Deviation σ_n	Sample Size n	Average $\bar{y}_n$	Standard Deviation σ_n
15------	0.5128	1.0206	50------	0.5485	1.1607
16------	0.5153	1.0301	51------	0.5489	1.1623
17------	0.5174	1.0384	52------	0.5493	1.1638
18------	0.5196	1.0471	53------	0.5497	1.1658
19------	0.5217	1.0558	54------	0.5501	1.1667
			55------	0.5504	1.1681
20------	0.5236	1.0628	56------	0.5508	1.1696
21------	0.5252	1.0696	57------	0.5511	1.1708
22------	0.5268	1.0754	58------	0.5515	1.1721
23------	0.5283	1.0811	59------	0.5518	1.1734
24------	0.5296	1.0864			
25------	0.5309	1.0915	60------	0.5521	1.1747
26------	0.5320	1.0961	61------	0.5524	1.1759
27------	0.5332	1.1004	62------	0.5527	1.1770
28------	0.5343	1.1047	63------	0.5530	1.1782
29------	0.5353	1.1086	64------	0.5533	1.1793
			65------	0.5535	1.1803
30------	0.5362	1.1124	66------	0.5538	1.1814
31------	0.5371	1.1159	67------	0.5540	1.1824
32------	0.5380	1.1193	68------	0.5543	1.1834
33------	0.5388	1.1226	69------	0.5545	1.1844
34------	0.5396	1.1255			
35------	0.5403	1.1285	70------	0.5548	1.1854
36------	0.5410	1.1313	71------	0.5550	1.1863
37------	0.5418	1.1339	72------	0.5552	1.1873
38------	0.5424	1.1363	73------	0.5555	1.1881
39------	0.5430	1.1388	74------	0.5557	1.1890
			75------	0.5559	1.1898
40------	0.5436	1.1413	76------	0.5561	1.1906
41------	0.5442	1.1436	77------	0.5563	1.1915
42------	0.5448	1.1458	78------	0.5565	1.1923
43------	0.5453	1.1480	79------	0.5567	1.1930
44------	0.5458	1.1499			
45------	0.5463	1.1519	80------	0.5569	1.1938
46------	0.5468	1.1538	90------	0.5586	1.2007
47------	0.5473	1.1557	100-----	0.5600	1.20649
48------	0.5477	1.1574	150-----	0.5646	1.22534
49------	0.5481	1.1590	200-----	0.5672	1.23598
			-----	0.5772	1.28255

TABLE IV

DETERMINATION OF THREE POINTS ON REGRESSION LINE

Let $m/n + 1$ equal*	Then y equals*
0.20	-0.47588
0.50	+0.36651
0.95	+2.97020

*These same values of $m/n + 1$ and y can be used for all problems, thus eliminating the necessity for computing or looking up values of y corresponding to various values of $m/n + 1$.

TABLE V

COMPUTATION OF CONTROL CURVES

When $m/n + 1$ equals*	Constants for $m/n + 1$ equals*
0.15	1.255
0.30	1.268
0.50	1.443
0.70	1.835
0.80	2.241
0.85	2.585
Point on line corresponding to:	
Second highest point on graph	0.75409
Highest point on graph	1.14071

*These values do not need to change from problem to problem.

CONTROL OF QUALITY OF SHEET METAL MATERIALS FOR CONTAINERS

S. J. Hotchner,* R. D. McKirahan,** and C. E. Lachele*
Technical Service Division, American Can Company

Like all container manufacturers, the American Can Company has the responsibility of supplying containers which will give adequate performance when used in accordance with the normal practices of the canning industry. Control of the quality of both the sheet metal used in the manufacture of metal containers and its in-plant processing is necessary to insure that the finished containers meet the manufacturer's quality standards.

The Canco program for controlling the quality of tin plate depends upon the cooperative efforts of the Manufacturing Department, the Technical Service Division, and the tin plate suppliers, with direction and coordination by the Purchasing Department. Deficiencies noted on receipt or in processing are reported directly to the supplier's representative by the can manufacturing plant. Manufacturing plants also participate in the testing of tin plate made to new specifications which may affect its fabrication.

Canco's Technical Service Division, which operates on a national basis, has been assigned the responsibility by the Purchasing Department for establishing specifications for all of the characteristics of tin plate associated with its performance requirements, both in the manufacture of containers and in their use by the canning industry. For this purpose, close cooperation is maintained between the various groups in each of the Technical Service laboratories as well as with the Manufacturing Department. The introduction of new organic coatings, new container designs, and new canning techniques at times impose different performance requirements for the tin plate. These are anticipated by coordination of the laboratory effort. Statistical samplings of tin plate deliveries are made as deemed necessary to follow the quality characteristics of one or more attributes, or to explore thoroughly any indications of improved or lowered quality of an individual supplier's deliveries.

Tin plate suppliers have a continuing program for improving the quality of their product. The Technical Service Division makes an important contribution to the quality improvement by bringing to each supplier's attention any observed deviations or areas where improvements are needed to provide for more demanding performance requirements. Meetings of suppliers with the Purchasing Department and the Technical Service Division promote the development and thorough evaluation of changes in materials and manufacturing methods with the assurance that the desired results are being obtained.

Basic Container Material

The basic material of the metal container industry is a low carbon steel which is commonly produced in one of four chemical compositions,

* Western Laboratory, San Francisco, Calif.
** Central Laboratory, Maywood, Illinois

depending upon the corrosion resistance and strength requirements. It is available in eleven different gauges, commonly referred to as basis weights, ranging from 75 lb. (0.0084 in.) to 118 lb. (0.0130 in.)* Both lighter and heavier basis weights than these are used for special applications. The plate is produced in tempers ranging from T-1 which is "dead soft" to T-6, the hardest available material. The corresponding R-30T values are 49±3 and 70±3. The particular combination of basis weight and temper is usually determined by the can size and the strength requirements.

A small amount of untinned steel is used for non-corrosive products such as dry cocoa and motor oil. For these applications the steel is enameled to afford protection against rusting. However, by far the largest portion of steel used by the container industry is coated with tin to provide the necessary corrosion resistance. Tinning is accomplished by either of two methods: hot dipping for the heaviest coatings, and electroplating for the lighter coatings. Electrolytic plate is produced both as cut sheet and in coils; hot dipped plate is available in cut sheets only. The hot dipped plate is supplied in two coating weights for most uses today; electrolytic tin plate comes in basically four different coating weights. (1)

Quality Testing for Standard Materials

The steel mills are constantly testing the mechanical and chemical properties of the tin plate during production to make sure that it meets specifications. Statistically established sampling systems and AQL's are in use for specific tests. The American Can Company has arranged with all the steel mills to accumulate test data on characteristics important for adequate performance of the material produced for its orders. These data are reported to the Purchasing Department and Technical Service Division laboratories either monthly or quarterly, depending on the nature of the product. The mill reports are basic to our approach to incoming material quality. They serve as a record of the quality of the supplier's production.

The assurances provided by the mill reports make it unnecessary to resort to the usual acceptance sampling and quality control procedures. No acceptance and rejection tests are made, and no AQL is established. However, quality investigations have been made to supplement the supplier's testing. These provide information on the more important attributes which enable one supplier's material to be rated against the others. Trends toward both improved and lowered quality can be detected so that appropriate correction can be made.

Statistical sampling of coil stock is performed at the American Can Company coil centers; mill cut sheets are sampled at the can manufacturing plants. Both the sampling program and the quality tests are the responsibility of the Technical Service Division. Reports in graphical form with appropriate comments are prepared at regular intervals for discussion at meetings of the suppliers with the Purchasing Department and the Technical Service Division.

* Basis weights and tin coating weights are expressed in units of "pounds per base box". A base box is the equivalent area of 112 sheets of plate 14 in. by 20 in. or 31,360 square inches.

QUALITY INVESTIGATION

TIN COATING WEIGHT #25 ELECTROLYTIC

7/1/57 - 6/30/58

(3 Spot Sheet Average)

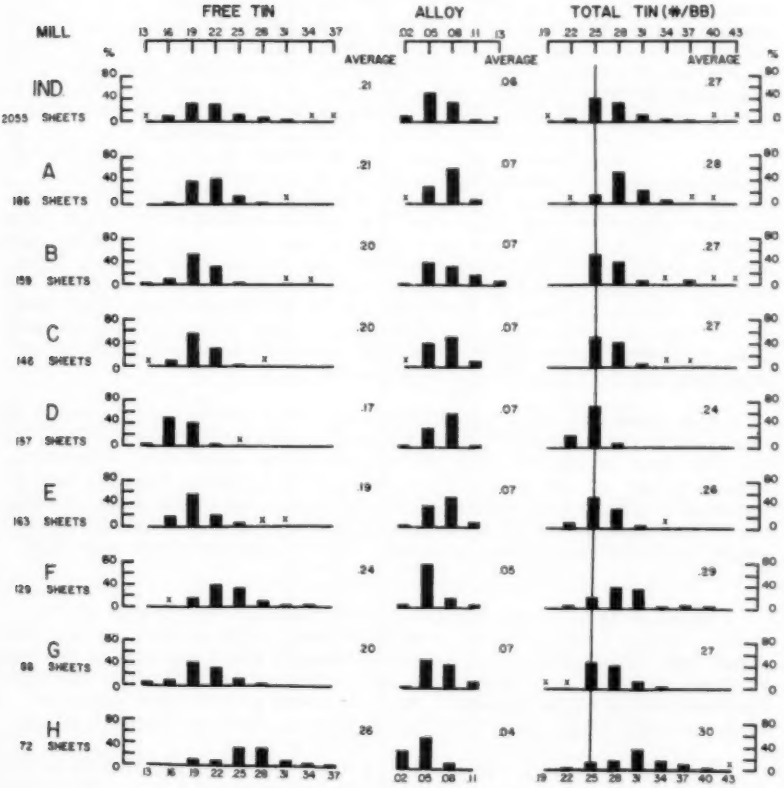


CHART I

The three major quality considerations with standard materials are discussed separately below, with examples of quality problems.

Mechanical Properties – This classification includes the specified characteristics such as tin coating weight, basis weight, temper, sheet size, and count per unit. These characteristics can be measured by simple objective tests and lend themselves to effective control by the usual quality control procedures during production at the mill. Results from Technical Service quality investigations over the years have shown excellent conformance with the values established by the American Iron and Steel Institute. (4)

Two quality problems of a more technical nature deserve some note here. Chart I is a portion of the 1957-58 annual summary of quality investigation testing of tin coating on No.25 electrolytic plate (0.25 pounds of tin applied per base box). These data were obtained by using an analytical technique which permits rapid and accurate measurement of the amount of both free tin and alloy tin. (2) Both are reported on the chart, together with the total tin coating weight – the sum of the two.

Until this method of analysis was developed in 1951, there was no convenient means of measuring tin in the alloy layer on electrolytic plate. With the initial use of this method in the 1952 quality investigation program it was found that a sizeable proportion of the No.25 electrolytic plate had relatively high amounts of tin in the alloy layer. Samples from one supplier averaged 0.10 lbs./BB tin in the alloy layer and the industry production averaged 0.07 lbs./BB. This fact was of concern since with more of the total tin in the alloy layer, less free tin is available for corrosion resistance. High alloy layers also are associated with soldering difficulties and with reduced brightness, an important trait of the tin plate container.

Tin plate suppliers were given these facts and asked to control the amount of alloy on their plates, a factor heretofore generally ignored. The 1957 data on Chart I show the results of the suppliers' efforts: average weight of tin in the alloy dropped to 0.06 lbs./BB, and more important, none of the individual averages were more than 0.07 lbs./BB. All of the average total tin coating weights were within the accepted range, and only two sheets out of the 1753 tested were less than 0.20 lbs./BB.

Hardness is a rapid, easily measured index of the temper of the steel. Temper U plate is produced from steel strip treated in the relatively new (1950) continuous annealing process which imparts to the steel a unique combination of mechanical properties in comparison to the traditional box annealed process. Chart II shows a five-year record of hardness tests from the quality investigation tests on Temper U plate.

In 1953, a sizeable proportion of the Temper U plate tested in the quality investigation program fell below the expected range. This fact was brought to the attention of the suppliers. Improvements were made in their processing procedures which resulted in a striking improvement in the hardness test results in 1954 and following years. With the hardness problem under control, continuously annealed plate has come into its own. About a dozen continuously annealed plate lines are now in

QUALITY INVESTIGATION

% R30-T TESTS BELOW RANGE ON T-U

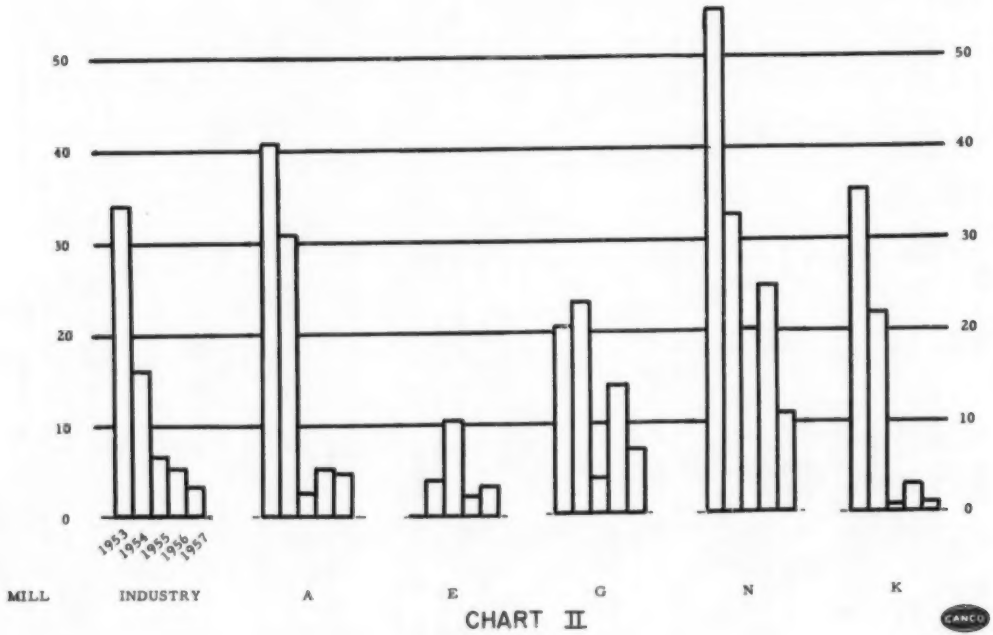


CHART II

CANCO

operation and six more are being installed or are in the advanced planning stages.

2. Surface Properties — Quality characteristics in this classification are generally not of the type that can be measured readily by objective tests. Canco depends on its suppliers for consistent, uniform processing and an understanding of the performance requirements. Producers have been given a series of tests to evaluate the surface characteristics of tin plate. These involve the use of specially formulated enamels to measure adhesion, flow-out and similar properties. They are used at the mills as quality control procedures and afford a measure of the plate quality for enameling and lithographing. There is a high probability that plate which passes these rapid evaluation tests will enamel and lithograph satisfactorily in our operations.

3. Corrosion Resistance — Over the past 30 years much has been learned of the factors affecting the corrosion of tin plate. The role of steel composition was well defined by 1940 and firm specifications established. The influence of steel making practices was well understood by 1950. Intensive work on tin coating requirements was begun at the start of World War II and is still under study today. In the last decade a major break-through on the control of corrosion resistance was realized with the development of "special properties" tests. These permitted the successful introduction in 1951 of No. 100-25 differentially coated electrolytic plate.*

a. Ladle Analyses

Control of corrosion resistance of tin plate starts with the steel refining process where the chemical composition is adjusted by the selection of scrap and open hearth additions. A heat of steel varies between 100 and 500 tons, depending upon the size of the open hearth. In the new basic oxygen process 50 to 100 tons of steel are produced at one time. A sample from the heat for chemical analysis is obtained from the ladle as the steel is being poured into the molds.

Type MR steel is the base metal most commonly used by can manufacturers. Adequate resistance to internal can corrosion is controlled by limiting the phosphorus, sulfur, and silicon contents. (3) Careful control of carbon and manganese also is important to assure steel with the required physical and metallurgical properties.

Chart III is a typical supplier's report on the ladle analyses of Type MR steel heats produced during the second quarter of 1958.

b. No. 100-25 Tin Plate

The suppliers' reports for No. 100-25 electrolytic tin plate include data for tin coating weights and the "special properties" tests (pickle lag, iron solution value, and tin crystal size). The latter tests are measures of the corrosion resistance and are used by the suppliers for control purposes to assist in obtaining and maintaining the

* No. 100-25 tin plate has an inside surface coated with the equivalent of 1.00 lbs. of tin per base box and an outside surface coated with the equivalent of 0.25 lbs. of tin per base box.

TYPE MR LADLE ANALYSES

SECOND QUARTER 1958

Carbon	No. Tests	%	Manganese	No. Tests	%	Phosphorous	No. Tests	%
.04 or Less			.20			.005 or Less	1	.5
.05			.21 - .25	-	-	.006 - .010	133	62.4
.06	5	2.3	.26 - .30	4	1.9	.011 - .015	77	36.2
.07	35	16.4	.31 - .35	29	13.6	.016 - .020	2	.9
.08	46	21.6	.36 - .40	39	18.4			
.09	13	6.1	.41 - .45	22	10.3			
.10	31	14.6	.46 - .50	41	19.2			
.11	50	23.5	.51 - .55	65	30.5			
.12	29	13.6	.56 - .60	13	6.1			
.13	4	1.9						

Sulphur	No. Tests	%	Silicon	No. Tests	%	Copper	No. Tests	%
.020 or Less	24	11.3	.004 or Less			.01 - .02	1	.5
.021 - .030	154	72.3	.005			.03 - .04	85	39.9
.031 - .040	35	16.4	.006			.05 - .06	94	44.1
.041 - .050			.007			.07 - .08	22	10.3
.051 - .060			.008			.09 - .10	6	2.8
			.009			.11 - .12	4	1.9
			.010			.13 - .14	1	.5
			.010 or Less	213	100.0			

CHART III

processing conditions consistent with the test requirements.

"Pickle lag" is a measure of surface properties which are primarily influenced by the hot strip and annealing operations. Iron Solution Value (ISV) is thought to be primarily influenced by the efficiency of the alkaline cleaning and pre-plating acid pickling. (5) Our present aims are for pickle lags of 10 seconds maximum and ISV values of 20 micrograms of iron maximum.

In 1957 the size of the crystals in the tin coating of electrolytic plate was shown to have an important influence on corrosion resistance. Large crystals are desirable.* Plate with small size tin crystals is diverted to non-critical applications.

The extent to which the No. 100-25 plate produced by each mill meets the special properties test requirements is a controlling factor in determining which mills can be considered acceptable sources. The status of each mill is reviewed quarterly on the basis of the Technical Service Division test results.

Chart IV shows a typical supplier's monthly report of quality tests on No. 100-25 tin plate. Tight control over the actual amount of tin coating is possible on this material, as illustrated by the distributions at the top of the chart. The range of coating weight determinations is within ± 0.10 lbs. tin/BB which is in accord with the American Iron and Steel Institute values (4). By way of contrast, the hot dipped plate grade designated as Common Cokes, averages 1.10 lbs./BB and individual tests will range from below 0.80 lbs./BB to above 1.50 lbs./BB, a spread of over ± 0.30 lbs./BB.

Quality Evaluation of New Materials

No material quality program is complete unless there are provisions for the smooth transition from standard to improved or new materials. This is particularly true in these days of accelerated advances in technology and economic pressures. The demands of increased competitive situations and for increased production efficiency must be met without sacrificing container quality.

Certain factors essential to satisfactory container formability and performance are not yet readily measured by objective tests. This is particularly true in the area of surface factors which affect enameling and lithographing characteristics. In the development of a new or improved treatment, the mill applies the special tests developed to assess the quality of standard materials. This is considered a partial evaluation and is supplemented by more exacting tests at one of our Technical Service laboratories.

The Technical Service Division, on receipt of information of a promising change from the supplier's technical department, arranges through the Purchasing Department to obtain samples for evaluation. These

* The ASTM standards for grain size of non-ferrous metals at a magnification of 1X is used for classification of tin crystal size. Large crystals are defined as those of size ASTM ± 9 or larger.

SUMMARY OF TIN COATING, RATE OF PICKLE LAG, IRON SOLUTION VALUES AND TIN CRYSTAL SIZE ON 1.00/.25# CONTINUOUS ANNEALED TINPLATE PRODUCED FOR AMERICAN CAN COMPANY DURING OCTOBER 1958.

No. of Sheets Tested - 39

TIN COATING

	EDGES		CENTERS	
	No.	Acc. %	No.	Acc. %
.92 and under			1	2.56
.93- .97	3	3.85	18	48.71
.98-1.02	50	67.95	18	94.86
1.03-1.07	19	92.31	2	100.00
1.08- and over	6	100.00	0	
	78		39	

No. of Sheets Tested - 39

PICKLE LAG

0-5 sec.	63	80.77	33	84.62
6-10 sec.	15	100.00	6	100.00
	78		39	

No. of Sheets Tested - 47

ISV

0-10 micro Fe	28	59.57	29	61.70
11-20 micro Fe	17	95.74	17	97.87
21-80 micro Fe	2	100.00	1	100.00
	47		47	

No. of Sheets Tested - 47

TIN CRYSTAL

ASTM #7 or larger	55	58.51	32	68.00
#8	24	89.04	15	100.00
#9	11	95.74		
#10	4	100.00		
	94		47	

Note: Coils having a tin crystal size of #10 or smaller on orders specifying minimum crystal size were held for American Can Company disposition.

CHART IV

samples range in size, depending on the nature of the change, from a few sheets to a cross-section of a full week's production. The product resulting from such a large test run is diverted to non-critical uses. On recommendation from Technical Service, the Purchasing Department advises the supplier if results of the evaluation warrant approval. Further, initial receipts of commercial deliveries are closely followed through production and in the field when deemed advisable. In this manner, improved surfaces for enameling, reflecting both chemical treatments and oils, have been developed and adapted as standard materials with a minimum of disruption to our manufacturing operations.

New materials, as opposed to new surface treatments, present a similar problem in their introduction, but the changes are usually of a more radical nature and more extensive testing is required. Suppliers may be asked to submit additional quality reports over those already established for standard materials. Until the material is accepted as a standard item, and this may take several years, samples selected at stated intervals during production at the mill are submitted to the Technical Service Division for rigorous study. Every aspect of a new material is evaluated in this way before the material is considered a standard item.

With fundamental changes in steel making practice or tin coating methods, evaluation of corrosion resistance is the primary purpose. A recent publication of our laboratories (6) gives details of the material quality problems that were involved during the period of transition from hot dipped to No. 100-25 plate.

Taking Action

A program such as has been described is only complete when there are effective means to implement it. The information collected must be put to work; if it is not, the quality program is non-productive.

To avoid this pitfall, a close supplier-buyer relationship is a necessity. This is effected by frequent visits by local representatives of the suppliers at our manufacturing plants and at the Technical Service laboratories in addition to the normal Purchasing Department contacts. Deviations in laboratory spot checks and progress in the evaluation of new and improved materials are pointed out as they occur. In this way the supplier is kept directly informed at the operating level, and delays in transmittal of such information are avoided.

An important link in the Canco-steel industry relationship is a series of meetings in which the more important aspects of tin plate quality are discussed. These meetings are held about every three years, and are attended by representatives of our Purchasing, Research and Technical, and Manufacturing Departments and representatives from the supplier's Sales, Research and Technical, and Production Departments. The presence of the top management of both the supplier and our company insures that there is over-all understanding of our observations and objectives.

Data from our tests are used to show quality trends and the position of the supplier in relation to competition. These data are also compared with the supplier's information on the same properties. Out of this phase of the discussions come both commendations for jobs well done and suggestions for improvements. The hazards and complications likely to be

encountered with use of material out of specifications are reviewed so that our problems are well defined to the suppliers.

Summary

American Can Company's quality program for sheet metal materials involves close cooperation with the suppliers. Periodic reports are furnished by the steel mills giving accumulated quality control test data on characteristics that are important for adequate performance of the material. These reports show the quality of the plate produced for our orders. The Canco laboratories make random spot checks of delivered material to discover any trends toward changes in quality. To assure suitability of new or improved materials for their intended use, there is intensive joint evaluation with the suppliers on the effect of critical change in plate manufacturing methods. In this way the quality of new or improved materials is brought under control before large scale commercial applications are effected. Examples of these procedures have been discussed. Contacts are maintained with the suppliers on all management, technical and operating levels to provide for the complete and frequent exchange of information essential to the implementation of the quality program.

REFERENCES

1. Hoare, W. E., Tinplate Handbook, 3rd edition, Tin Research Institute, Greenford, Middlesex, England. October 1956.

2. Kunze, C. T. and Willey, A. R., Electrolytic Determination of Tin and Tin-Alloy Coating Weights on Tin Plate, Journal of Electrochemical Society, 99, No. 9, 354, September 1952.

3. Hartwell, R. R., Certain Aspects of Internal Corrosion Resistance in Tin Plate Containers, Advances in Food Research, Vol. III, Academic Press, New York, 1951.

4. Tin Mill Products Section, Steel Products Manual, American Iron & Steel Institute, New York, October, 1957.

5. Willey, A. R., Krickl, J. L., and Hartwell, R. R., Steel Surface Properties Affect Internal Corrosion Performance of Tin Plate Containers, Corrosion, 12, No. 9, 433, September 1956.

6. McKirahan, R. D., Connell, J. C., and Hotchner, S. J., Application of Differentially Coated Tin Plate for Food Containers, Food Technology, 13, No. 4, April, 1959.

PANEL - "PROCESS CONTROL OF CONTINUOUSLY EXTENDED MATERIALS"
PAPER PROCESS CONTROL

Frederick C. Hartwell
Quality Control Coordinator
Peter J. Schweitzer Division
Kimberly-Clark Corporation

I can hardly do justice to either the quality control profession or the paper industry in the ten minutes that I have been alloted by our Moderator. However, I do appreciate the opportunity of being here and will attempt to give you a few highlights covering the process control of paper. In my alloted time I would like to list briefly the unit operations involved in making paper and mention quality control's participation and then spend a few minutes showing examples of what I believe are a few basic but important techniques used in setting up process control and reject specifications in the mill.

The paper process includes:

(a) raw material acceptance and preparation - chemicals, supplies, pulps
(b) cooking and bleaching - wood, flax, rag, waste
(c) beating and refining - pulp preparation
(d) stock regulation and dilution - changing from batch to a continuous process
(e) sheet formation - laying stock out onto a moving wire
(f) water removal - leaving paper fibers formed together while removing excess water
(g) drying - getting rid of most of the water
(h) windup - making a jumbo roll of paper
(i) finishing - slitting, sheeting and fabricating units for shipment

Quality control plays an important part in the above-listed paper-making process. Quality control functions, both statistical and non-statistical, include:

(a) raw material control - establishing and maintaining sampling test and inspection procedures, setting material specifications and obtaining certificates of compliance on incoming materials.
(b) process control - developing and maintaining sampling, testing and inspection procedures, establishing mill manufacturing control limits and performing process engineering studies.
(c) production control - developing and maintaining sampling test and inspection procedures

for the finished product, establishing audit systems, vendor contact, issuing certificates of compliance, evaluating customer complaints and evaluating the competitors' product.

Although some techniques of statistical quality control differ in application between industries, most are applicable to the paper industry. Certainly, control charts, ($\bar{X}$, R, C and p charts) are used. Sampling plans, analysis of variance, t and F tests, correlations, non-parametric tests, etc. are all statistical tools being used today in the paper field. It is not my intent to cover these, but to show you only one or two applications of basic statistical techniques that might stimulate your thinking along similar lines in your respective industries.

It has been my experience that mill control limits for process control (as distinct from mill reject limits) are extremely helpful to manufacturing groups as a means of keeping a process at desired quality levels. It is important when establishing these limits that all values be realistic in terms of customer needs and process capabilities. Some properties are more difficult to control than others; some are dependent on each other and all vary in their relative importance.

Slide 1 - My first slide shows an example of the use of probability paper in helping set realistic manufacturing control limits based on past performance data. The curve shown here is developed by making a histogram of values accumulated over six months of data for property A, and by using standard techniques[1] the data are plotted on probability paper thus transforming it into easily workable information. With data plotted in this manner, we can estimate the percentage of paper testing below or above a certain value. For example, property A on Slide 1 shows 5% less than 1600 and 50% of the values below 1875. If we want to assume that we can do as well or better in the future, as we have in the past, we can then use this curve to establish machine control limits for property A. For example, assuming that future performance is no worse than past performance, limits of 1540 to 2210 will result in no more than 5% outside of these limits. (Quite often, limits are used such that they include 90% of past performance values; in this case 1600 to 2150).

If we want to write reject specifications for customers we will often use 95%-99% limits.

By watching long term trends (Slide 2) we may constantly upgrade both manufacturing control limits and customer reject specifications.

Since the manufacturing control limits obtained in the above method include considerable assignable variation (that variation which can take place within six months of manufacture), it is sometimes wise to calculate close tolerance limits on those properties that the operator can control at the paper machine; for example, basis weight. In this case, we want the machine personnel to aim for target and to keep the basis weight running down "the middle of the road" so to speak. Since this is usually a property that can be controlled directly by the

machine-operating personnel, it is possible to give him narrow limits; yet not so narrow that he is adjusting the process when he should not. One technique used in calculating these limits is to measure machine direction and cross machine direction variation in basis weight using operator and testing equipment actually used for the process control and then by analysis of variance, strip out the machine direction and cross machine direction variation and calculate the residual, or inherent, variance. This residual variance is essentially test error. From the residual variance, upper and lower limits for basis weight control can be calculated. With limits calculated in this manner, machine personnel are more apt to make adjustments only when a significant change in process level has taken place.

Limits for the range of weight from maximum to minimum across the machine may readily be calculated by using this residual variance and the appropriate d_2 factor(2).

Slide 3 is an example of a control chart form for plotting machine direction and cross machine direction using X and R values.

Slide 4 shows a thickness control chart for use in plotting machine direction and cross machine direction, $\bar{X}$, R values for both front and back rolls off a paper machine. Similar techniques for calculating residual thickness variance are used as for basis weight residual variance.

Slide 5 shows the manufacturing control limit form which is posted at the paper machine. Limits posted represent realistic values determined statistically and based on either minimum variability studies (such as for basis weight) or on values calculated from past performance data. The specification limits also listed on the form would normally be customer reject specs. Of course, customers often fix these specifications, but with studies similar to those described above, the producer is usually in an excellent position to persuade the customer as to where the specifications should be set.

There are many other applications in paper process control for statistics. For example:

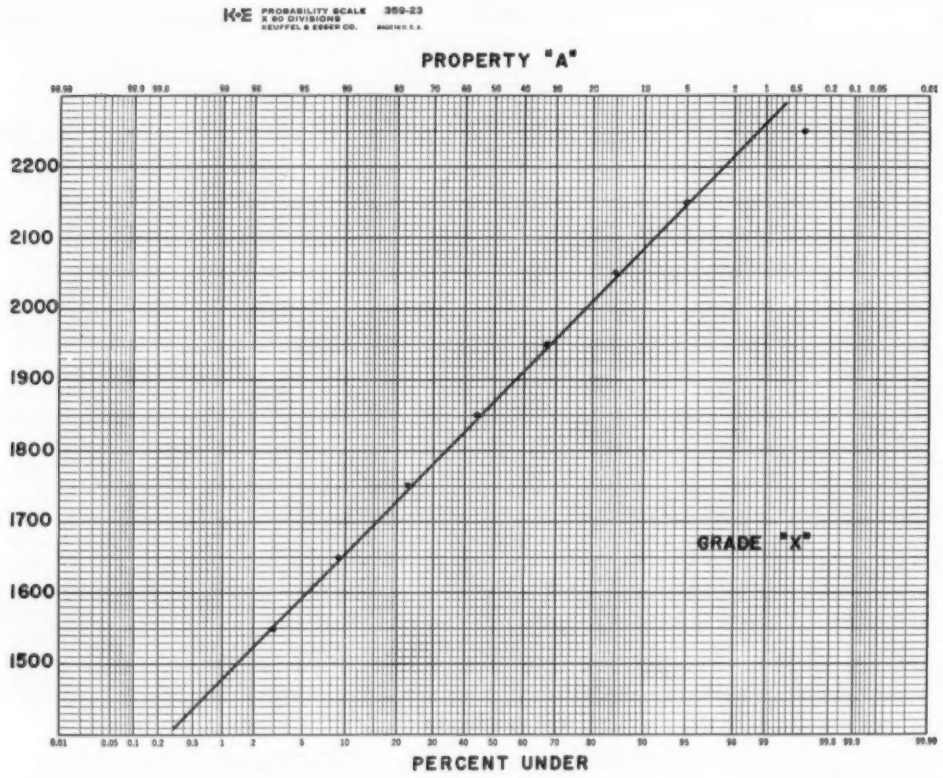
(a) to compare differences between operators, testers and methods
(b) to set up proper sampling techniques
(c) to evaluate process changes
(d) to set up useful audit systems
(e) to install quality indices
(f) to minimize trim, losses
(g) to optimize yield
(h) to sample work loads
(i) to increase the efficiency of experimentation.

Certainly those of us working in statistical quality control in the paper

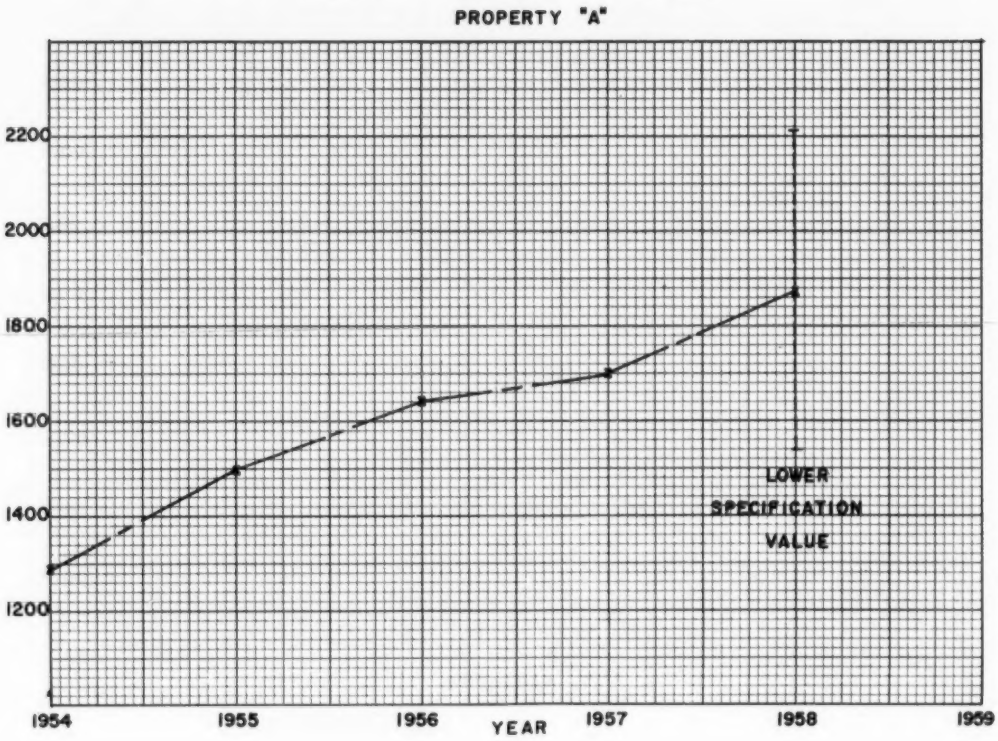
industry recognize that we are in virgin territory. We expect to see many new uses of this managerial tool developed during the next decade.

References: (1) "A Graphical Method of Determining the Characteristics of a Frequency Distribution", by Arthur Bender, Jr. Delco Remy Div. GMC

(2) "Quality Control and Industrial Statistics", by A. J. Duncan, Richard D. Irwin, Inc. (1952)


K-E PROBABILITY SCALE 359-23
X 90 DIVISIONS
KEUFFEL & ESSER CO.
PROPERTY "A"
2200
2100
2000
1900
1800
1700
1600
1500
GRADE "X"
0.01 0.05 0.1 0.2 0.5 1 2 5 10 20 30 40 50 60 70 80 90 95 98 99 99.8 99.9 99.99
PERCENT UNDER

10 X 10 DIETZGEN GRAPH PAPER
10 X 10 PER INCH
EUGENE DIETZGEN CO
MADE IN U. S. A.

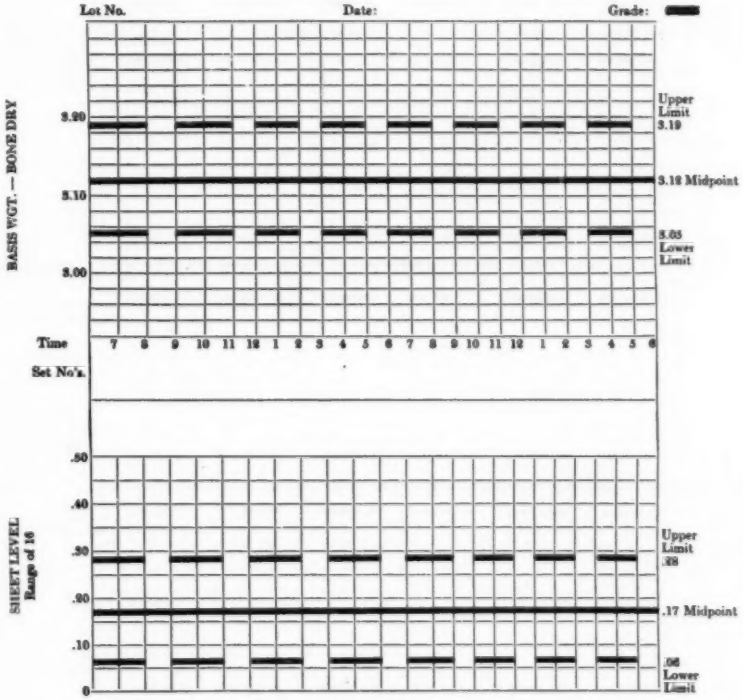


Form 10018

MACHINE CONTROL CHARTS
BASIS WEIGHT

[redacted]

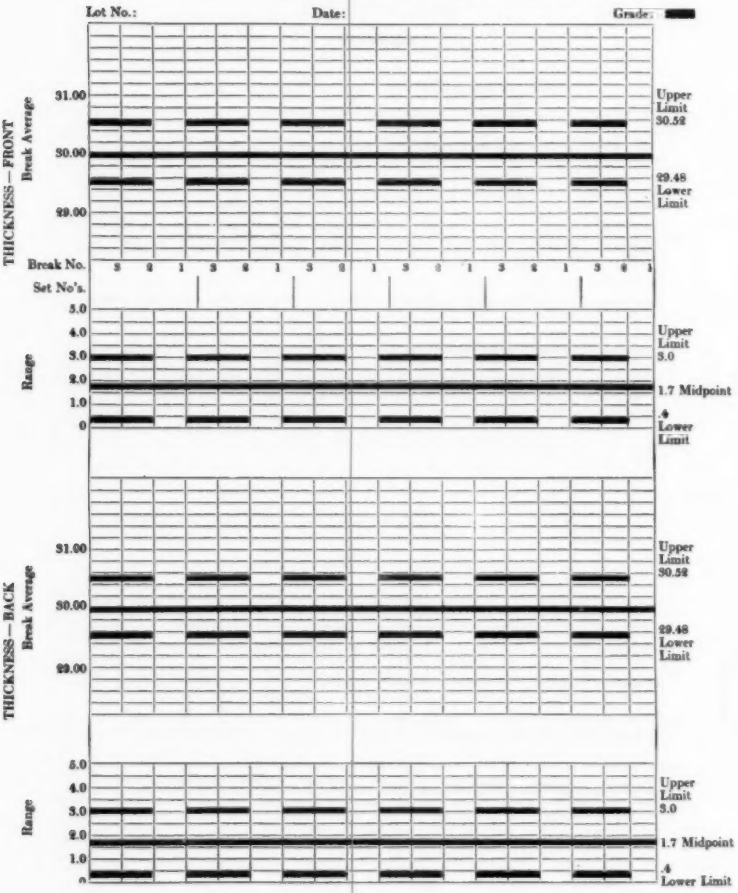
CONTROL LIMITS BASED ON 3 SIGMA' (RESIDUAL VARIATION)



Form 10010

SUPERCALENDER — THICKNESS CONTROL CHARTS

CONTROL LIMITS BASED ON 3 SIGMA' (RESIDUAL VARIATION)



Machine Control Limits and Specification Limits

Grade _____ Customer _____

Mfg. Control Limits	Properties	Inspection Reject Limits
	A	
	B	
	C	
	D	
	E	
	F	
	G	
	H	
	etc.	

Special Information:

Date Issued _____

Date Revised _____

QUALITY CONTROL APPLICATIONS IN THE OIL REFINING FIELD

David Frazier and D. R. Whitman*
Research Department
The Standard Oil Company (Ohio)

Our subject, "Quality Control Applications in the Oil Refining Field", is easily disposed of: There aren't any. This doesn't mean that oil companies aren't concerned with quality or its control. They are, and they have people specifically charged with "quality control", but the words mean something different to them than they do to the great majority of members of the American Society for Quality Control. In the petroleum industry "quality control" means a sort of auditing or accounting function, assuring that what we make meets specifications, but without any considerations of randomness or probabilities: talk of "X-bar" and "R" charts, standard deviations, three sigma limits, consumer's risk, average outgoing quality limit, or sequential sampling, for example, just isn't heard in petroleum company offices.

These concepts, like many others of modern statistical quality control, are useful and important, and the fact that the petroleum industry doesn't use them doesn't mean the industry is backward or lazy. The reason is more fundamental than that: modern "quality control" methods were developed to apply to production of discrete things, bobby pins, ball bearings, bomb sights, or H-bombs. But in the petroleum industry we don't make discrete things, we make continuously flowing materials: gasoline, motor oil, fuel oil, wax. And production of such continuous materials differs from that of "things" in two very important ways:

1. There is no natural "unit" of continuous production. If the things we're talking about are ball bearings, we can discuss 100 bearings or ten or one, but "one tenth of a bearing" is meaningless. If we cut up a bearing it isn't a bearing any more. Not so with continuous materials like gasoline: we can meaningfully talk about a property of ten gallons of gas, or one gallon or one tenth or one hundredth, or any other quantity large or small.
2. Continuous production can be mixed--we call it "batch averaged"--with consequent averaging out of fluctuations it may contain. For example, if we have a barrel of motor oil half of which is too thick and half of which is too thin, it is entirely possible that by simple mixing we can make the whole barrel acceptable. But a box of ball bearings half of which are too big and half of which are too small is forever unacceptable.

It's these differences that make conventional statistical quality control of little applicability in the oil refining industry. What we're going to do here is suggest a system that is applicable and show that this system is nothing more than a simple and natural generalization of conventional quality control.

*Present address: Assistant Professor of Chemistry, Case Institute of Technology, Cleveland 6, Ohio.

Now let's consider two highly hypothetical industrial plants, one making ball bearings, and the other making lubricating oil. We're going to talk about just one property of the bearings, their diameter, and just one property of the lubricating oil, its viscosity. (Viscosity measures how easily a liquid flows: water has relatively low viscosity, molasses has a much higher viscosity.) And let's not be distracted by the facts that real ball bearings are never quite perfectly spherical, and that viscosities don't usually average in a simple way, etc. This is just an illustration, not a practical application.

Now consider that ball bearing plant: It has an excellent quality control engineer, and by astute use of modern quality control methods and a lot of old-fashioned horse sense, he has production in an excellent "state of control". What does that mean? He could probably be induced to agree with this mathematical definition:

$$P_i = A + E_i \qquad (1)$$

where P_i represents the diameter of any arbitrarily designated bearing, A is the target value, and E_i is an error randomly and independently drawn from a normal population with zero mean and a satisfactorily small standard deviation, σ_ϵ. Most of the technical terms here are quite familiar to quality control engineers and need not be stressed; they are used in their regular senses. But it might be well to review what that "independently" means, since it is at the heart of the troubles that are experienced when anyone tries to apply conventional quality control to continuous production.

Here "independently" means that if I tell you what A, the target value is, and then require you to guess what was the diameter of a particular bearing, it won't help a bit if I also tell you what the actual measured diameters were of, say, the previous and following bearings: you'd just guess the target in any case.

This assumption of independence, particularly serial independence, is important, and it's involved in one way or another in most quality control tables. To be sure it's justified we test a conventional "X-bar" chart by examining the number of "runs" in a series of readings. If there are too many or too few we conclude the process is "out of control" and we look for the cause.

We can get away with this assumption in the case of ball bearings because of their discreteness. I can examine every thousandth bearing, or four every hundred, or even every bearing, but no one will ask me to test every one-tenth of a bearing: that's meaningless. Thus, to go back to equation (1), I can assume that this equation holds "for all i".

But now consider the lubricating oil plant. Feedstock is coming in through a pipe under the fence, and the product is going out through a pipe under the same fence. I can't tell the errand boy to go out and pick up "four oils out of every hundred". Instead I have to tell him something like, "Go get me a quart sample every hour", or "Take a five

gallon sample from each ten thousand gallons". But, and here's the important point, there's no reason why I can't get a sample every hour, or minute, or second, or tenth of a second. It still would be recognizably lubricating oil, but I simply must believe that as the samples get closer and closer together, the time will come when the assumption of independence would break down. That is, I'll find that a sample taken now is much more like the one taken one second ago than it is like the one taken an hour or a day ago; what I know about mixing of liquids in tanks and pipes tells me that much.

Thus the general assumption of independence among samples is untenable in continuous production, and the usual tables and quantitative rules are useless. What can we do about this? Well, let's recognize that in production of continuously flowing materials, the sample I take now is always to some extent contaminated with the material that was in the unit when I took the previous sample. Let's adopt this equation as our ideal "state of control":

$$P_i = \rho P_{(i-1)} + (1-\rho)(A + E_i) \qquad (2)$$

Here as in Equation (1) A is the target value, E_i is an error randomly and independently drawn from a population with zero mean and satisfactorily small standard deviation, and P_i is the result of any arbitrarily designated analysis on the stream. There are two new terms here: $P_{(i-1)}$, which is the result of the previous analysis, and ρ (the Greek letter rho) which measures the extent to which the present sample is contaminated with material which was in the unit when the previous sample was taken. It will always thus be a number between zero and one. If it is zero equation (2) becomes the same as equation (1) and conventional quality control would apply.

Equation (2), we suggest, should be adopted as the mathematical model of "state of control" for continuously flowing materials. It's intuitively attractive, and lots can be done with it: it turns out that this is an equation well known in mathematics. It defines what is called a "linear autoregressive Markovian process", and lots is known about it and hence about any production unit which is acceptably "in control" in this sense. Here are two of the more important properties of series generated by this equation:

1. The P's have a well defined central value. There is no tendency to wander off farther and farther as time passes.
2. There is no periodicity in the usual sense. If the process is "up" now it doesn't tell me that it will be "up" again at some future time.

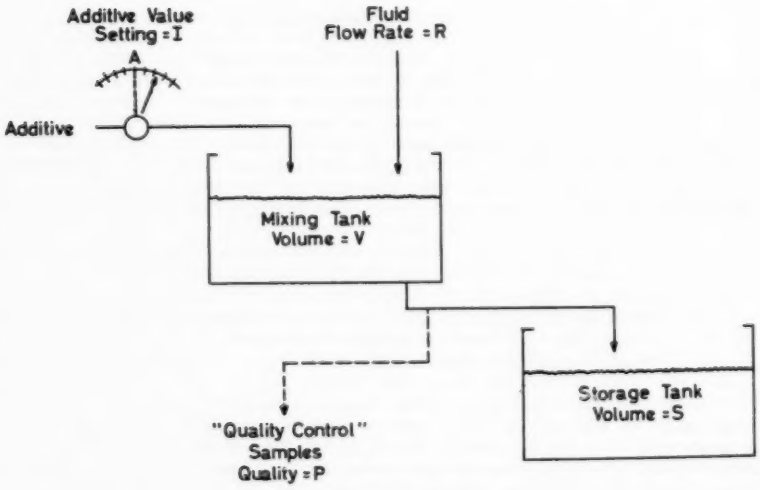
Another nice thing about this equation is that it can be "realized" by the interesting, though highly idealized, physical process pictured in Figure 1. Think of this as a process in which a large volume of liquid is being treated with a small amount of additive. Examples might be coloring of gasoline or chlorination of water. The fluid to be treated flows at a rate R into a mixing tank of volume V and on into a storage, or batch, tank of volume S. All

three of these quantities will be thought of as being perfectly constant, and mixing in the processing tank will be thought of as perfect. The additive coming in from the left will be thought of as being so concentrated that its volume is negligible. This concentrate flows through a control valve whose intended setting is A, the target value, but which actually fluctuates around A. Now suppose these fluctuations take place regularly and very frequently in such a way that the departures are randomly, independently, and normally distributed about the target. If the output stream is sampled at regular intervals much longer than the valve fluctuation period, it can be shown that the sample qualities will follow this relation:

$$P_i = \rho P_{(i-1)} + (1-\rho)(A + E_i) \qquad (3)$$

This is, of course, exactly the same as equation (2). Thus, the physical process illustrated in Figure 1 would be in a "state of control" according to the definition given in equation (2), which as we have seen, can be related to a "state of control" for conventional discrete production. Equations (1) and (2), and the model Figure 1, are the bridge by which we will bring quality control of continuous production, so typical of the petroleum industry, into the main body of "statistical quality control".

FIGURE 1



Equation (2) and the model of Figure 1 have twin properties that make them useful here:

1. The physical system they suggest is intuitively, subjectively attractive: it "looks a little like" a real petroleum or chemical process.

2. The equations and definitions are sufficiently exact that we can do interesting things with them mathematically.

Now look again at equation (3). It was written in this particular form to make it look as much like equation (1) as possible; that is, it was written this way to emphasize the similarity between our suggested state of control for continuous processes and the familiar state of control for discrete processes. Having made that point, let's rewrite it in another form, recognizing that actually we can't count discrete units of production, by substituting t, the time at which the sample was taken, for i, the sample number, and substituting Δ (delta), the time since the last sample, for the 1 in the expression (i-1). We have:

$$P_{(t)} = \rho P_{(t-\Delta)} + (1-\rho)(A + E_{(t)}) \qquad (4)$$

If samples are taken regularly at intervals of Δ , it can be shown that:

$$\rho = e^{-\frac{R}{V}\Delta} \qquad (5)$$

Here e is the base of natural logarithms, R is the flow rate, V is the mixing tank volume, and Δ is the sampling interval. Equation (4) thus becomes:

$$P_{(t)} = e^{-\frac{R}{V}\Delta} P_{(t-\Delta)} + (1 - e^{-\frac{R}{V}\Delta})(A + E_{(t)}) \qquad (6)$$

If the "additive valve" in Figure 1, which is thought of as undergoing random, independent, and normal fluctuations about A, is further thought of as undergoing these fluctuations very rapidly with respect to the sampling interval Δ , then the errors of equation (6) ("$E_{(t)}$" terms) can be shown to be randomly, independently, and normally distributed with standard deviation given by:

$$\sigma_E = \sqrt{\frac{K}{\Delta}} \qquad (7)$$

Equations (6) and (7) have only three unknown parameters. They are:

A, the target value, set by the operator.
K, which measures the inherent error or variability of the system.
R/V, which measures the resistance of the system to change.

Compare these with those of equation (1), which idealizes conventional quality control. It has two parameters:

A, the target value, set by the operator.

σ_E, which measures the variability of the system.

Our model has these, and adds only one more, the measure of the resistance of the system to change.

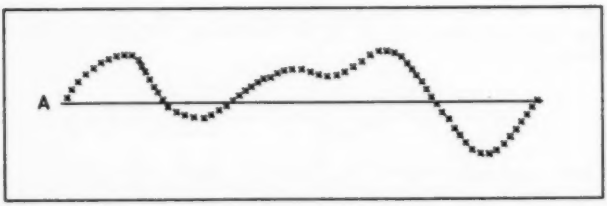
Applications of the Model

Suppose we have a continuous process producing, say, gasoline, and let us say that somehow it has been brought into a reasonable approximation to a "state of control" as we define it. Later we will have some suggestions as to how this might be done, but for the moment let's assume it has been done. Of course our model won't fit the real situation exactly: no model ever fits the real world exactly. But suppose it fits tolerably well. What can we do with it?

1. Quality Control Charts

Every now and then someone decides he's going to apply modern quality control to a petroleum stream; of course he begins by plotting a quality control chart, and that's where the matter ends, too, because the chart looks something like this:

FIGURE 2
QUALITY CONTROL CHART



There's nothing much you can do with this except to watch for obvious changes in level. It fails the usual tests for trends, and runs, and almost anything else, and there's usually nothing you can do to the process to make it produce a chart of the kind you find in any quality control textbook. This process is just "out of control" by usual standards.

But it might be in control by our definition. Consider again equation (4): Solve it for $(A + E_{(t)})$:

$$(A + E_{(t)}) = \frac{P_{(t)} - \rho P_{(t-\Delta)}}{(1 - \rho)} \qquad (8)$$

Even if the consecutive readings, the P's, are strongly correlated (that is, ρ is large) this equation will in effect remove the correlation. For example, if the chart of Figure 2 were from a real "in control" process, and if somehow we had a good estimate of ρ, equation (8) applied to the points pair by pair would give a series of independent estimates of $A + E_i$ which when plotted up would look like any well behaved discrete process quality control chart. It would be possible to apply the usual tests for runs, etc., and action limits and rules could be set up in the usual way.

2. Changes in Sampling Interval

For an "in control" discrete process nothing much happens if the sampling interval is changed. However, in continuous production we know that as the sampling interval is decreased, the samples must become more alike. This effect can be estimated quantitatively by equation (5):

$$\rho = e^{-\frac{R}{V}\Delta}$$

As an example of the use of this relation we will observe that when the sampling interval, Δ becomes so short that ρ becomes as large as about 0.9, there is little to be gained in sampling oftener. On the other hand, if ρ is less than about 0.1, the process is in effect "running by itself" for an uncomfortable fraction of the time.

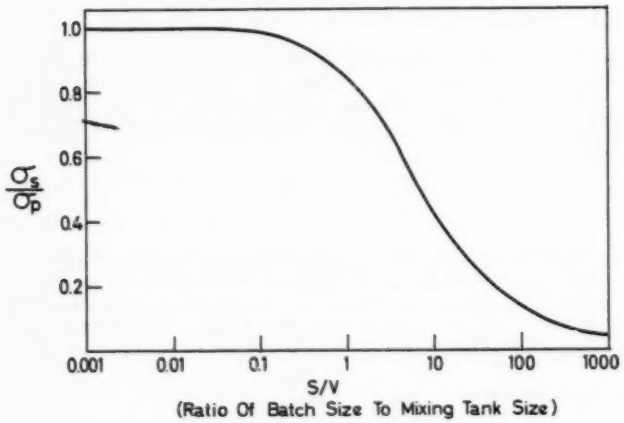
3. Batch Variability

A striking difference, previously mentioned, between continuous and discrete production is that in the former quality fluctuations ordinarily can be averaged out, and that the bigger the batch, the more such averaging there is. For example, if the hypothetical ball bearing producer we were talking about a while ago makes bearings which are, on the average, one per cent defective, he will find that one per cent figure holds whether he ships his product in boxes of one hundred or one million or ten. But a petroleum refiner who is shipping million gallon batches of gasoline, and who is controlling quality so that only one per cent of his batches are below specification in some critical property, may well find that if he decreases batch size the batch rejection rate may go up for this reason alone. It is therefore of interest to know how batch variability may be affected by batch size for our model of continuous production. It turns out to be relatively easy to figure this, and results are presented graphically in Figure 3.

Here the horizontal axis is the batch size in terms of V, the apparent volume of the hypothetical mixing tank illustrated in Figure 1. The vertical axis is the ratio of the standard deviation of batches of the specified size to the standard deviation of small "spot" samples of the same stream. An illustration may make this clearer. Suppose we have a lubricating oil stream and by a method we will discuss later we have estimated that the effective mixing tank volume of the process producing the stream is 1000 gallons. Also, we have been taking samples of the stream for quite a while and we find that the standard deviation of their viscosities is about five units. Suppose we are

going to ship this oil in 10,000 gallon tank cars. What may we expect the standard deviation of average tank car viscosities to be? Here the batch size 10,000/1,000 or ten. Entering Figure 3 at batch size of ten, we find the standard deviation ratio is about 0.4. Therefore we estimate the batch standard deviation to be about 0.4 times 5, or two viscosity units.

FIGURE 3
BATCH STANDARD DEVIATION



On the other hand, if this product were to be shipped in fifty gallon drums, batch size would be 50/1,000 or 0.05, for which we find the standard deviation ratio to be about 1.0. Therefore the standard deviation of drum size batches of this product will be about five viscosity units.

Suppose management policy was that not more than one batch in a hundred could be below viscosity specification. How much above specification must the operator shoot? From standard tables we find the one per cent level of the single tailed "t" to be about 2.33. Multiplying this by the standard deviations above, we find that the operator must shoot about 4.7 units high for tank car production, or about 11.6 units for drum production.

4. Batch Quality Estimation

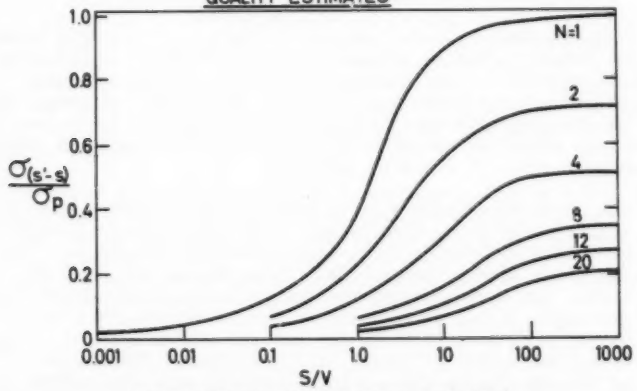
A very common practice in the petroleum industry is to collect intermediate products in quite large batches. After they are collected, they are mixed and then one or more samples is taken. Future processing depends upon the outcome of the analyses. Automobile gasoline is usually made this way. Batches of a million gallons or more of nearly finished fuel are collected. They are, however, deliberately below specification in certain key properties, most notably in

"octane number". (Octane number measures resistance of a fuel to the undesirable phenomenon variously called "knock", or "carbon knock", or "pinging".) To bring them up to "anti-knock" specification a small but extremely important volume of tetra-ethyl lead (sometimes called "Ethyl fluid") is added.

This lead is expensive. To minimize its use, each batch of semi-finished fuel is well mixed, sampled, and analyzed, and only just enough lead is added to bring it up to specification. In principle this works all right, but in practice it's not entirely satisfactory because it takes a long time to mix such a vast volume, and after that it takes several hours to get the sample, analyze it, and get the results back to the operator.

However, it is usual to take samples of the stream at regular intervals during the production of the batch. It would save a lot of time and effort if these could be used to estimate the batch quality, and it turns out this can be done. There is some uncertainty in such an estimate, of course, because we don't know exactly what the stream may have been doing between samples, but if it's in control by our definition this uncertainty can be measured and allowed for. The derived relationship is expressed graphically in Figure 4. The horizontal axis is the batch size in terms of the theoretical mixing tank volume, as in Figure 3. The vertical axis this time is the standard deviation of difference between the estimate of batch quality and the true batch quality, in terms of the long term standard deviation of samples. The estimate of batch quality is just the average of all readings taken during its production. (Readings are assumed to be equally and centrally spaced.) Again, an example: Suppose we are making 1,000,000 gallon batches of semi-finished gasoline from a process which has an effective mixing tank volume of 10,000 gallons, and suppose we have found from experience that the long term standard

FIGURE 4
STANDARD DEVIATION OF BATCH QUALITY ESTIMATES



deviation of samples taken from this stream is 0.7 anti-knock (octane) units. In the course of each batch's production we are going to take and analyze four samples, and we'll use the average of these four samples as our estimate of the true batch quality for figuring how much lead to add. How far off may we expect this average of four readings to be? Here batch size is 100; entering the graph there and going up to N=4 we find the standard deviation ratio to be about 0.5. Therefore the standard deviation of the estimation error will be 0.7 x 0.5 = 0.35 numbers, and since lead addition can be made very accurately, this 0.35 number will be the standard deviation of actual outgoing batches. Suppose management has said that only one batch in a hundred, on the average, may be below specification. As in the previous example we see that the operator must shoot 2.33 x 0.35 = 0.82 numbers high.

Shooting this high would be very expensive; however, using the curves of Figure 4 we could find out how many tests we'd have to do to reduce the overshoot costto acceptable levels. If we could put a dollar cost on each test we could even figure the optimum number of tests; in fact this was the problem that started this whole research.

Practical Considerations

We have put off until now two related awkward questions:

1. How do we know when a process is "in control" by our suggested definition?
2. If it is in control, how do we get the values necessary for the suggested computations?

To begin with, we know that no real process will fit our definition exactly, and certainly some won't fit it at all. But some processes fit to a useful degree, and others, we suspect, could easily and profitably be made to fit to this degree. What constitutes a useful fit, and how a non-conforming process might be made to fit, will have to be up to the skill and judgment of the quality control engineer in each particular case.

There are several mathematical and statistical tools that might be used in studying a series of readings from a process to estimate how close it might be to our definition. For practical purposes, however, we suggest the following procedure which lacks mathematical elegance but has the advantage of ease of application. It is intended for situations where the stream's natural variability is large compared to the inherent error of a single analysis.

1. Get some "long" series of analyses of the stream. The analyses should be equally spaced in time, and by "long" we mean there should be at least twenty or thirty crossings of the center line. Plot them up and look at them. Are there obvious changes in level or any clear cycling? If so, the process is not in control, and the causes will have to be sought out and corrected.
2. Count the number of center line crossings. The number of them should be quite a lot less than half the number of points. That is, if we have hourly readings for a

week, 168 readings, the number of center line crossings should be much less than 84, probably nearer twenty or thirty. If the number of line crossings is near half the number of points, the sampling interval is so long that the methods we've talked about here are not applicable.

3. When you have some series that pass the two tests above, adopt a trial value for R/V, the ratio of flow rate to apparent mixing tank volume. Sometimes the actual hold-up of the apparatus can be used for V. More commonly, however, the hold-up is much less than the apparent mixing volume because of controller characteristics, thermal inertia, or process kinetics. Using the trial value for R/V and the known sampling interval, Δ, calculate a trial value of ρ using equation (5). Plug this into equation (8) and recalculate the whole series point by point. Plot the new series of $(A + E_4)$ values. If you have about the right value of R/V, and hence of ρ, the number of center line crossings will be about half the total number of points, and the series will look like a conventional quality control chart, and it can be treated as one. If there are too few center line crossings, make the tank bigger, and if there are too many, make it smaller.

If you can find a value of ρ which seems to give series which are "in control" by ordinary quality control engineer's standards, assume that the parent process is "in control". Then go back and figure the standard deviation of a long series of analyses. This number and the apparent mixing tank volume obtained above are the only numbers needed to apply the methods we've been talking about.

Precautions

The method given here for finding the apparent mixing tank volume by trial and error is quite inexact. It is much better to estimate it by calculating the "serial correlation coefficient for lag one". This is just the correlation coefficient formed by pairing the first reading with the second, the second with the third, the third with the fourth, and so on to the end of the series. Statistical textbooks, for example M. G. Kendall's Advanced Theory of Statistics, give examples of this calculation.

Everything given here has been on the assumption that the analytical error is small compared to the process variability. This is often the case in the petroleum industry, but it is by no means always the case. When it is not, modified formulas and graphs must be used which take the analytical error into specific account. Mathematical details of this work will be published elsewhere.

Summary

Conventional statistical quality control methods aren't applicable to the continuous processes characteristic of the petroleum industry

because in such processes quality readings must be "serially correlated", and "batch averaging" must take place. However, a model for "state of control" has been suggested which is applicable, and a physical realization of this model has been suggested. Using this model makes these things possible:

1. A quality control chart with conventional properties can be constructed for a continuous process.
2. Effect of changing sampling interval in such a process can be estimated.
3. Variability of batch averages about the target can be estimated.
4. Variability of quality estimates derived from sample averages about true batch qualities can be estimated.

QUALITY CONTROL ENGINEERING IN PROCESS AND PRODUCT CONTROL

Robert Schin
Western Electric Company

Those of you who have attended the previous sessions of this series on professional quality control engineering work are by now substantially aware of the wide scope of this function. A complete program of total quality control for a manufacturing company might well encompass the engineering use of statistical tools in basic product design; evaluation and control of purchased material quality; investigation, analysis and control of manufacturing processes; and rating of quality and reliability of completed product. A wide variety of sampling plans for lot by lot acceptance, rejection or quality rating may be used extensively from incoming materials and parts, within plant processing operations, and on through to completed product. However, the measurement, analysis and eventual economic control of variation in a manufacturing process through the use of Dr. Walter A. Shewhart's control chart theory, is today recognized as one of the most important functions of the quality control engineering profession.

The Western Electric Company, with its objective of high quality at minimum cost, has placed considerable emphasis on the control of its own manufacturing processes with the time-honored philosophy, "that you cannot inspect quality into a product, it must be built into it". The effectiveness of this emphasis on process control has provided feed-back information for improved product design, reduced inspection effort and considerably reduced scrap, rework and sorting. Needless to say, all of these have contributed substantial savings in manufacturing costs. For example, at the Allentown Works of Western Electric, a plant of 3,000 people, where we manufacture electronic devices, tangible dollar savings have exceeded 20 million dollars since 1952. Western is continually expanding the application of these principles and techniques throughout its many manufacturing locations to optimize economic product quality and reliability. The end result is demonstrated by the outstanding performance and service cost to the users of the world's largest and most complex machine, the Bell Telephone System. This complex communication network has been essentially manufactured by the Western Electric Company.

While we do have a total quality control program in Western Electric, which requires the services of many Quality Control (QC) Engineers of professional calibre, I shall describe only the in-process quality control engineering function as it is carried out at the Allentown Works. This job of engineering in-process controls is comprised of seven major facets. (Chart 1)

QUALITY CONTROL PROGRAM

1. TRAIN ENGINEERING AND SHOP PERSONNEL IN THE UNDERSTANDING AND USE OF STATISTICAL TECHNIQUES AND CONTROLS

2. MAKE CAPABILITY STUDIES OF EACH STEP IN THE PROCESS BY:

   A. COLLECTING DATA
   B. ANALYZING DATA
   C. INITIATING ACTION TO ELIMINATE CAUSES OF OUT-OF-CONTROL CONDITIONS

3. INSTALL PERMANENT SHOP CONTROLS TO MAINTAIN PROCESS AT OPTIMUM ECONOMIC CAPABILITY

4. PERIODICALLY REVIEW CONTROLS TO ADJUST TO CHANGING CONDITIONS

5. TROUBLE SHOOT OUT-OF-CONTROL CONDITIONS AFTER PROCESS CAPABILITY HAS BEEN ACHIEVED

6. EXPERIMENT WITH PRODUCT DESIGN, PROCESS CONDITIONS AND MANUFACTURING FACILITIES TO FURTHER IMPROVE PRODUCT QUALITY AND COST

7. EVALUATE AND REPORT PROGRESS TO MANAGEMENT

Chart 1.

QUALITY CONTROL TRAINING

ALLENTOWN WORKS

1952 - 1958

COURSE	HOURS	GIVEN TO	NO. OF PEOPLE	TOTAL HOURS
MANAGEMENT	10	STAFF	15	150
BASIC	18	SUPV. & ENGINEERS	698	12,564
REFRESHER	3	SUPV. & ENGINEERS	280	840
SHORT BASIC	6	KEY SHOP PERSONNEL	305	1,830
SAMPLING INSPECTION	10	SUPV. & INSPECTORS	36	360
DESIGN OF EXPERIMENTS	21	ENGINEERS	252	5,292
2 WEEKS AT UNIVERSITIES	80	ENGINEERS	20	1,600
		GRAND TOTAL		22,636

Chart 2.

The first step in any program is to promote understanding. This applies to top management as well as to all levels of supervision, engineering and operating people. To accomplish this, the quality control engineers have prepared formal training manuals, tailored for specific groups, and have conducted all the in-plant training classes. At the Allentown Works, since 1952, more than 1,000 people have received over 22,000 hours of instruction. (Chart 2)

Upper management groups are given appreciation courses to enable them to understand and evaluate the overall Quality Control Program. Shop supervisors and product engineers are given an 18-hour basic course. This course includes Statistical Quality Control (SQC) theory and application, in addition to such subjects as interpretation of control chart patterns, the value and results of a complete process control program, and how to instill and maintain quality consciousness. Each participant is given enough training to understand the results of the statistical techniques used and to relate these results to his own practical or theoretical knowledge of the product or process. Key shop personnel such as layout operators, machine setters and process checkers are given a six-hour course covering application of basic techniques to provide them with a working knowledge of the statistical tools involved. In order to maintain plant wide understanding, most courses are updated and repeated annually for new personnel. Supplementing this formal training, each QC Engineer carries on a continuous educational program in his day-to-day contacts.

A basic philosophy of the Western Electric Company is to hold each operating shop responsible for the quality and cost of products they produce. Therefore, it is vitally important for first-line shop supervisors and their key people to thoroughly understand and accept the process controls used on their job. They must be convinced of the value of the control chart to the extent that they use it daily for efficient control. The QC Engineer develops and installs these process controls and instructions for the shop to use in controlling their own processes.

The training function is considered by Western as extremely important to the success of a quality control program. The QC Engineer, to carry out this phase of his job, must have sufficient statistical knowledge to simplify SQC theory and practice for use of shop personnel as well as to answer all questions that might be asked. Equally important, he must be a good instructor to convey understanding and acceptance of statistical techniques to people totally unfamiliar with these concepts.

As part of the overall program, many of Western's QC Engineers have contributed material during the past three or four years for preparation of a company handbook which was published in 1958. This book

Chart 3.

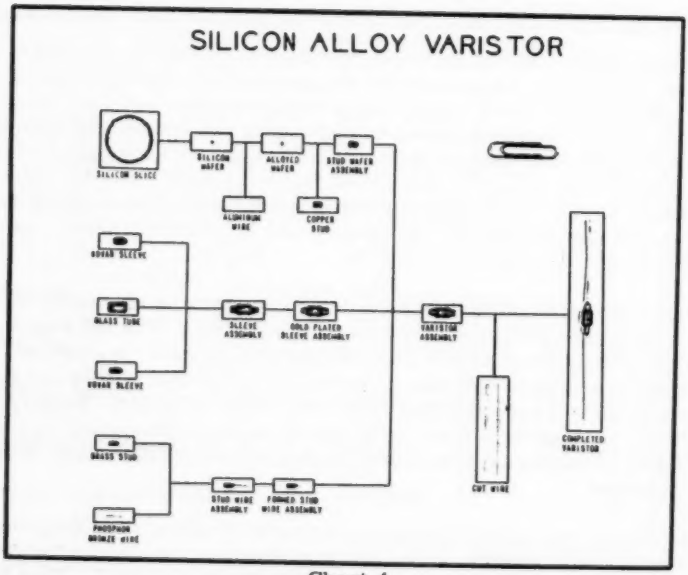


Chart 4.

records those elements of our quality control program which have been proven by actual practice to be economically sound. Every shop supervisor and product engineer has received a personal copy of this handbook which he is expected to use for self-education and reference. Each QC Engineer is expected to be completely familiar with this book to the extent that he can use and understand every principle and technique.

We have found that the installation and continuous operation of in-process controls can be most effectively carried out by "Quality Teams". (Chart 3) Each production line has its team, composed of the Shop Supervisor, Product Engineer and QC Engineer. Each member brings to the group his specialized knowledge. The Shop Supervisor brings practical knowledge. The Product Engineer brings product theory as well as equipment and process experience. The QC Engineer supplies both technical and statistical background. All three men on these teams are of equal stature in the eyes of the company. This provides a necessary atmosphere of mutual respect and cooperation. In practice, each QC Engineer is a member of a number of Quality Teams, covering several product lines. The Product Engineer and Shop Supervisor, on the other hand, are usually assigned to a single product line. The combined knowledge and exchange of ideas at weekly or monthly team meetings have been found to be the best method of accomplishing process control objectives.

The manner in which a Quality Team approaches the actual installation and constant improvement of process controls can perhaps be best illustrated with a specific example. I have selected as an example the Silicon Alloy Varistor, a product manufactured at the Allentown Works. This is a semi-conductor device, primarily intended for use as a rectifier and voltage limiter in telephone transmission equipment. While it is relatively simple in appearance, stringent electrical requirements make its manufacture a very exacting process.

Chart 4 shows a picture of a completed varistor and the components used in its assembly. On the top line, a slice cut from a silicon ingot is diced into 45-mil square wafers. The wafer, in turn, is alloyed with aluminum derived from the aluminum wire. The alloyed wafer is then soldered on to the copper stud, forming the stud wafer assembly. On the bottom line, a 10-mil phosphor bronze wire is inserted and soldered into a hole of the brass stud, thereby making the stud wire assembly. The wire portion of this assembly is formed into an "S" shape to insure a spring contact. These wire and wafer assemblies are force-fitted into opposite ends of the sleeve assembly, shown in the center line, which consists of two kovar sleeves fused to a multiform glass tube. Wire leads are then soldered to each end of the varistor, thus completing the assembly. The paper clip is shown for comparison of size.

Manufacture of this product was started in late 1955, and the process control program was initiated at the same time. At the first

SILICON ALLOY VARISTOR LOSSES

OPERATIONS	DECEMBER 1955 NET $ VALUE OF DEFECTIVES
CUT WIRE	14.84
WIRE STUD	22.57
COPPER PIN	10.28
WIRE STUD ASSEMBLY	128.41
SILICON WAFER	271.39
ALLOYED WAFER	926.13
ETCHED PIN-WAFER ASSEMBLY	2,959.31
ASSEMBLED VARISTOR	338.36
TESTED VARISTOR	8,213.45
TOTAL	12,850.44

Chart 5.

BREAKDOWN OF TEST DEFECTS

TYPE DEFECT	% OF DEFECTS
FORWARD VOLTAGE	5.3%
SLOPE	0.3%
REVERSE CURRENT	59.7%
OPEN	4.9%
SHORTS	10.6%
ZENER VOLTAGE	19.2%

Chart 6.

meeting of the Quality Team, the QC Engineer presented a breakdown of the cost of defectives (Chart 5) for a representative period. It was readily apparent that a high proportion of the defectives was dropping out at the testing operation. Attention was therefore directed to this item of greatest loss for the first process capability study. Further analysis showed that 60% of the testing defects were due to high reverse current (Chart 6). The team agreed to gather variables data and plot an $\overline{X}$ and R chart on this characteristic. The performance study made from January 3 to January 7 (Chart 7) showed out-of-control conditions on both average and range as indicated by the X's. Note that the grand average, or center of distribution is also above the maximum specification.

The QC Engineer has focused shop and engineering attention on the major problem, so that their skills and knowledge may be directed to an early solution. He now guides the interpretation and analysis of the data. From their knowledge of the product and processes, all members of the Quality Team suggest the factors at prior operations which they believe might be responsible for the erratic high reverse losses. (Chart 8) In silicon dicing, the wafer width was considered important; at alloying, temperature and belt speed were important factors; at wafer assembly, chips, cracks, off-center buttons and cleanliness; at pin assembly, alignment and shape of "S" bend. The team then decided to collect data and plot charts at each of these operations. Let's examine these charts to see the results obtained.

Initial variables data on wafer width (Chart 9) indicated a controlled process capable of meeting specification, but not properly centered. A series of dicing machine adjustments and operator instructions over a period of seven weeks eliminated all oversize wafers from the process.

At alloying, a percent defective chart (Chart 10) for visual defects indicated unstable conditions with about 30% of the units unsuitable for further processing. Sorting was introduced temporarily while the effects of lower furnace temperatures were studied. A reduction in temperature from 900° to 800° almost completely eliminated visual defects.

An electrical check made on the alloyed wafer (Chart 11) also showed process instability, averaging about 20% defective. After reducing furnace temperature to eliminate visual defects, the belt speed was adjusted to determine the effect on the electrical characteristic. Increased belt speed raised the level of defects. Finally, reducing the belt speed from 1-1/2" per minute to 1" per minute, decreased the electrical dropouts to an average level of about 2%.

At wafer assembly, where the alloyed wafer is soldered to one end of the stud, the skill of the operators was involved. Therefore, separate percent defective charts were plotted for each operator. Chart 12 shows the representative results of one operator. Off-center and cracked wafers were the major causes of erratic performance. Clarification

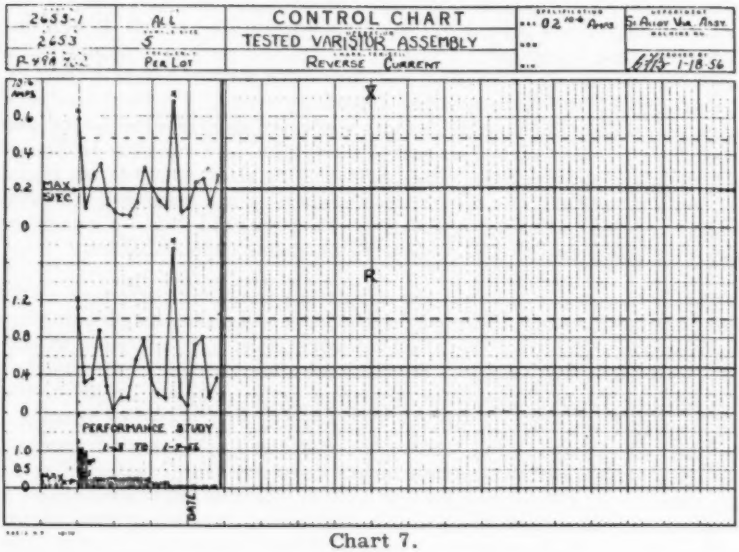


Chart 7.

FACTORS AFFECTING HIGH REVERSE REJECTS

PRIOR OPERATION	FACTOR
SILICON DICING	WAFER WIDTH
ALLOYING	TEMPERATURE BELT SPEED
WAFER ASSEMBLY	CHIPS CRACKS OFF CENTER BUTTONS CLEANLINESS
PIN ASSEMBLY	ALIGNMENT SHAPE OF "S" BEND

CONTROL CHARTS PLACED AT EACH OPERATION TO STUDY LEVEL AND VARIATION OF FACTORS.

Chart 8.

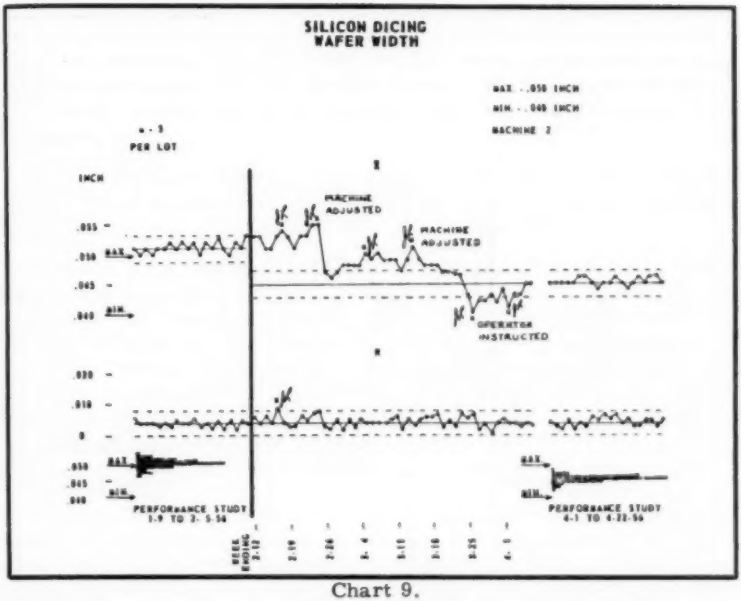


Chart 9.

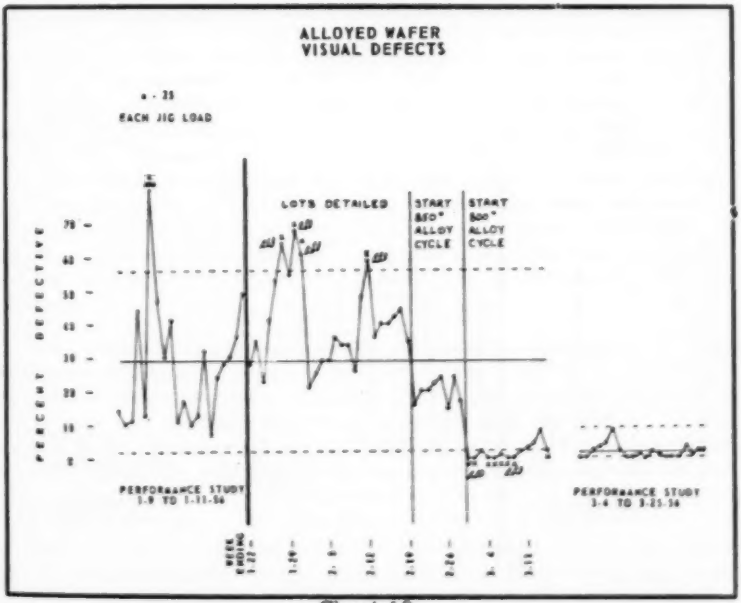


Chart 10.

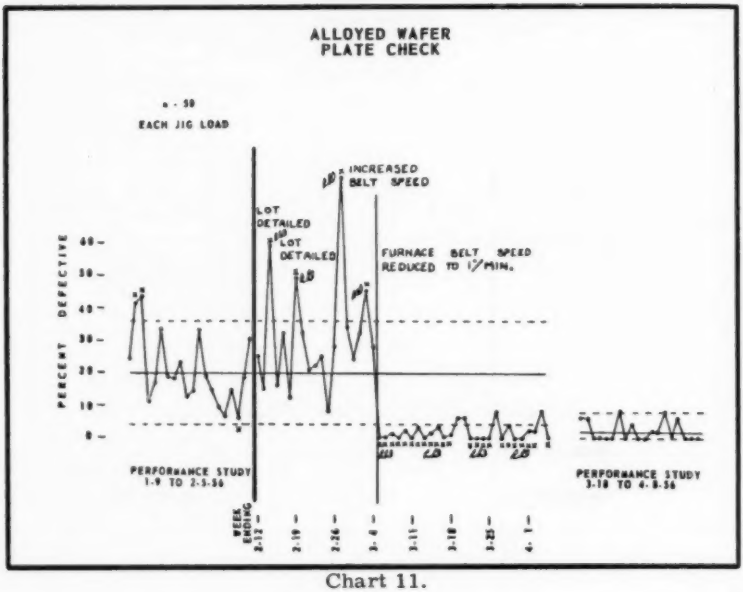


Chart 11.

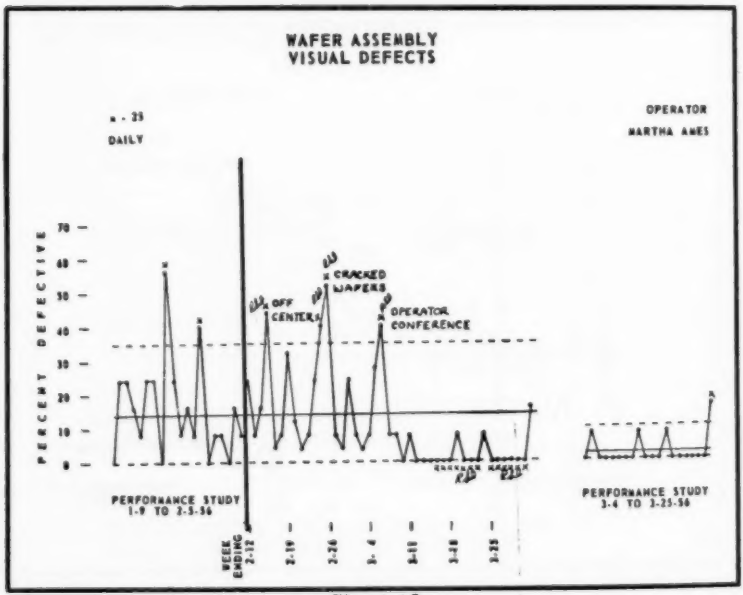


Chart 12.

of the assembly techniques, involving handling and location of wafers, as well as soldering temperature and time, were established through an operator conference with the Quality Team. Within one month after the initial study, visual defects at this operation were almost completely eliminated.

A simultaneous study on the formed pin assembly (Chart 13) resulted in modification of a forming tool which stabilized the operation and reduced repairs from about 35% to less than 10%.

In seeking improvement and control for one final electrical characteristic, the team found it necessary to investigate these earlier operations. Separate capability studies were necessary to bring each one into a state of control. By working back to basic causes affecting a process and instituting control of them, an overall process capability can be achieved. Coming back to the original objective to reduce electrical dropouts at final test, Chart 14 shows the results for reverse current after the earlier operations were stabilized. The percentage of defectives was reduced from 60% to 3% in approximately three months.

The Quality Team does not cease to function when capabilities are attained. To maintain continuous control of the processes, shop charts and instructions must be planned and installed. These controls must be reviewed periodically and adjusted to changing shop conditions. Charts must be discontinued when no longer of any value for control. Checking intervals must be extended to minimize quality checking costs as greater process stability develops. New charts may also have to be added as other unstable process variables are uncovered. The team must meet on a regular schedule to review all charts and work together to track down and eliminate causes of out-of-control conditions which creep into the process. Changing materials, new operators, machine wear, or adjustments due to maintenance can upset the control of basic variables. After the capability of a manufacturing process has been achieved, we may still not be satisfied with the final results. With a controlled process, it is then possible to experiment with product design and processes to make further improvement through the use of statistically designed experiments.

This example has portrayed how we install and maintain an in-process quality control program. The Quality Team has been shown as the key agent to accomplish the objectives. The QC Engineer, as a member of this team, is an integral part of the manufacturing group. To stimulate and guide the team activities, he must have a technical background to quickly and fully understand the manufacturing process and its problems. He must thoroughly understand and be able to properly apply the many SQC techniques. He must be the spark-plug and prime mover of the Quality Team, reaching ahead for improvement, initiating experiments, and guiding the team to new solutions of manufacturing problems.

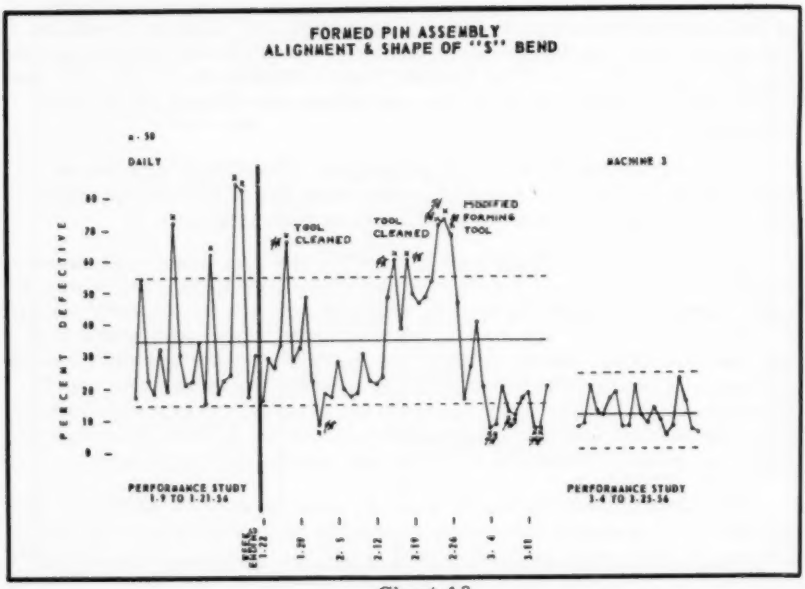


Chart 13.

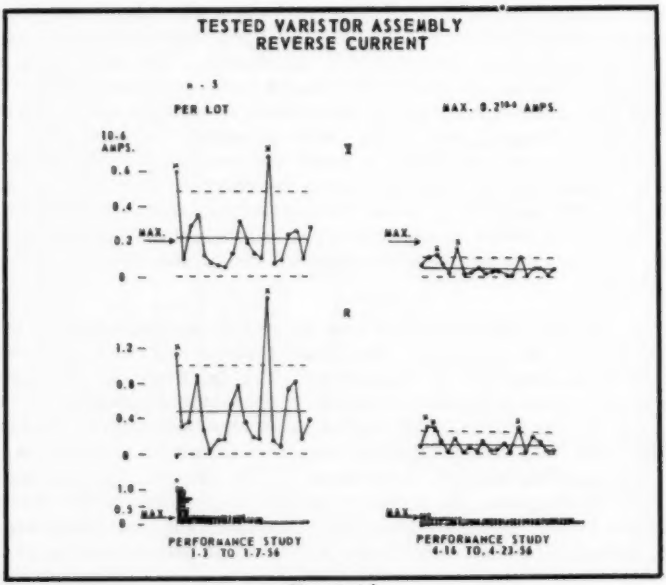


Chart 14.

It is essential for any functional group to satisfy its management that their work is contributing to the overall benefit of the company. The QC Engineer must likewise substantiate his efforts. One of several yardsticks used successfully at Allentown is the comparison of junk, repair and sorting costs, called Merchandise Losses, with the value of good product manufactured each month. For example, on the Silicon Alloy Varistor job, the reduction of merchandise losses, using December 1955 as a base period (Chart 15, Top), portrays the rapid progress made during the period when initial capability studies were being conducted. Much credit is also given to the Quality Team for the continuing stability and steady improvement shown up to the present time. The QC Engineer is required to forecast at the beginning of each year the improvement expected in merchandise losses by the end of the year. One measure of his performance is the attainment of these objectives. Reports and forecasts like this one for each product line are summarized into an overall chart for the Allentown Works (Chart 15, Bottom). These charts, of course, reflect total improvement which is attributable to other things as well as the use of SQC. However, of this total plant picture, almost 50% of the improvement since 1951, can be directly tied to the Quality Control Program. Probably some of this improvement would have been eventually achieved without SQC, but can we afford to wait for "osmosis" to take place.

Concurrent with cost reduction through process control, the quality of finished product must, of course, be rigidly maintained or improved. The Demerit Index, a quality rating to customer standards, is another yardstick by which management measures the effectiveness of its Quality Control Program. The Allentown Works' overall Index (Chart 16) shows consistent and steadily improving results. Mr. Patterson, of the Bell Telephone Laboratories, our next speaker, will tell you more about the Bell System Quality Rating Plan.

At the Allentown Works, over 500 different codes of product are manufactured each month on about 25 separate production lines. Over 4,000 different operations are being performed daily. Although only 30% of these operations vary sufficiently to warrant process controls, it requires over 5,000 control charts and 700 process control and inspection layouts to keep our manufacturing processes at top performance. New codes requiring new facilities and new processes are continually being introduced into manufacture. Time intervals from research to design to manufacture are being continually reduced. New products nearly always demand more stringent requirements. To keep pace with new science and new technology, we must quickly shake down, refine, and control our manufacturing operations. At Allentown and throughout Western Electric, the QC Engineer is playing an increasingly important role in the attainment of these objectives.

Our Quality Control Program was started and has been continued with graduate engineers from many engineering fields, such as electri-

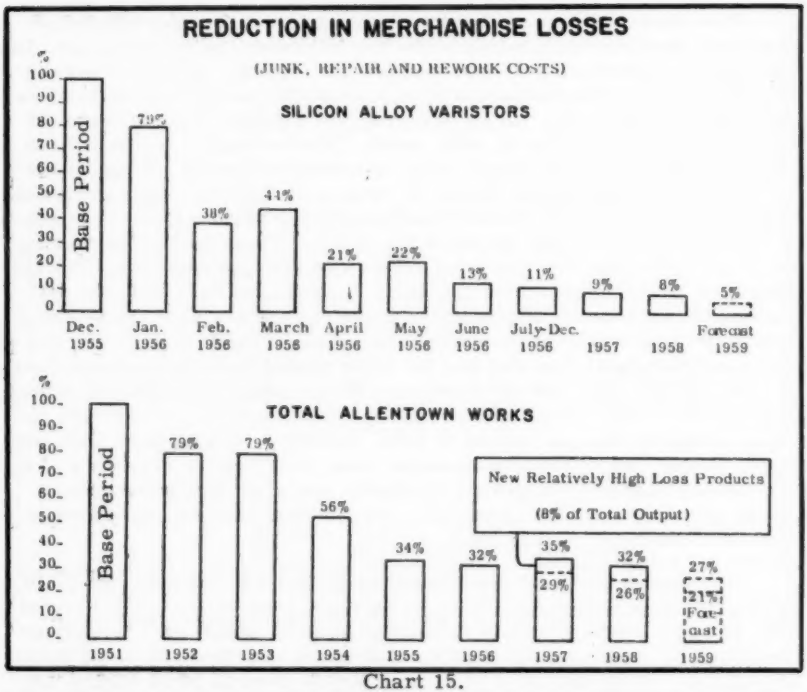


Chart 15.

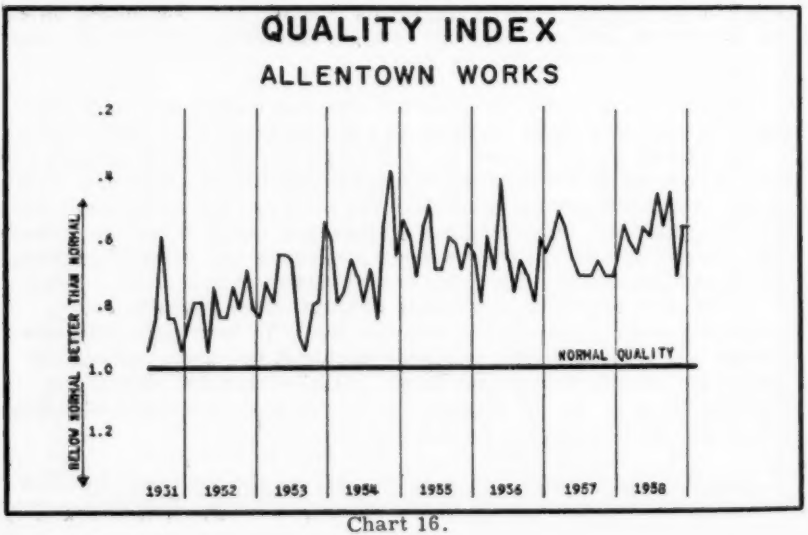


Chart 16.

cal, mechanical, chemical and industrial. They were given intensive training in SQC, both by the company and at various universities. I am looking forward to the day when a comprehensive statistical program will be included in an undergraduate engineering curriculum and a full-fledged quality control engineer is made available to industry. At present, only a few graduate schools, such as Rutgers, offer a full training program in engineering statistics.

Our program has been successful only because of the calibre and training of the men assigned to the job. I am convinced, and so is the Western Electric Company, that it is only by this combination of technical and statistical skill that a man can be equipped to meet today's quality control engineering challenge in industry.

Industry must recognize and accept the Quality Control Engineering profession and effectively utilize its services to attain greater productivity, better quality and lower costs.

USE OF OPERATIONS RESEARCH IN WESTERN ELECTRIC COMPANY

Robert E. Johnson, Economist and Actuary
Western Electric Company, Inc.

Last year I spoke to the Middle Atlantic section of this Society with regard to the broad spectrum of Operations Research applications to industrial problems. Since that time much water has flowed under the bridge and I am sure that most of you have seen at least some of the many articles pertaining to Operations Research that have been published during 1958. What I would like to do today is to discuss two specific applications recently completed by my group. Both cases involve basic Quality Control principles.

Of course, I have an ulterior motive in using these studies. That is, I would like to point out the tremendous need for statisticians and Quality Control people to exercise their imaginations in the areas of Operations Research. I strongly believe that if these techniques are to grow and the general usefulness of Operations Research is to expand, then this growth will come in the future equally through the intensified efforts of the applied statistician and the theoretical mathematician.

The first case that I would like to discuss is in the area of inventory control and production scheduling. Originally we were approached by one of our factory locations to develop required inventory levels for some 250 products in order that this plant might be able to operate most economically in a free market situation. This request came about as a result of several increases in the capacity of this location. Prior to these increases, demand had exceeded capacity, thus necessitating the allocation of products to the customers. In effect, this condition meant that practically everything produced could be sold.

Management was now concerned with establishing minimum inventory consistent with maintaining prompt deliveries under relatively unknown demand. Chart 1 shows the objectives of this program.

On the basis of these objectives, the following was developed (Chart 2). Eight groups were established predicated upon the level of demand in order to incorporate economic manufacturing runs. The column marked Production Cycle indicates the frequency with which each product would be manufactured, dependent upon the average incoming demand. The next two columns pertain directly to the mechanics of our control and scheduling plan. By control period we are referring to the time interval that is used in reviewing and determining action criteria. This will be made somewhat clearer in the next chart. The column marked Standard Deviation Factor is used as a criterion of demand volatility. These factors were developed from careful study of the few periods (although brief) in which some of the items were on a free market basis. These data were then extrapolated to cover all items.

Now, let us look at the mathematical formulation of the plan itself (Chart 3). Using a similar concept to the quality control chart ($\bar{X}$ + R), statistical limits were developed within which it was permissible for inventory to fluctuate (upper part of chart). When no significant change in demand occurred, production continued on a constant basis. When either the upper or lower control limit was pierced, action with

regard to production was required. This action is predicated on a servomechanism and the extent of remedial action is prescribed by the model as shown on the lower half of the chart.

Let us look at the first formula. This is used to get the new production rate when demand has lowered the inventory below the minimum desirable level. We wish to produce to the new demand rate on the next production cycle, and over a twelve-week period build the inventory back to the minimum level. K_rD then tells us the quantity to produce to meet the new demand level. The rest of the formula tells how much more to produce in order to replenish the inventory. $(k_rD + 2 Sk_rD)$ equals the minimum level of inventory we desire to obtain. However, in the interval before the new production volume is received, we may expect the present control value I to be augmented by production at the old rate A_o and to be decreased by the demand for the period until new production becomes available K_oD. Hence, the formula within the brackets $[\ ]$, tells how much inventory correction is needed while the fraction $\frac{k_r}{12}$ tells how much to produce in each future production run in order to achieve the desired inventory in three months.

Suppose we look at the bottom formula which demonstrates the technique in the case where demand has fallen and inventories are in a potentially overstocked condition. Again k_rD represents the amount we wish to produce to meet the new demand level. The remainder of the formula determines how fast we shall absorb the existing overstocks by reducing production below current demand. The first parenthesis ( ) within the first bracket $[\ ]$ repeats the anticipated changes in inventory before the new production level becomes operative. The second parenthesis within the bracket immediately reduces inventory to an upper action limit. The fraction, $\frac{k_r}{4}$, will achieve this adjustment within one month. The remainder of the equation in the second bracket plus its preceding fraction, $\frac{k_r}{12}$, is the same as we have seen in the first equation, but since it is preceded by a minus (-) sign, is designed to move the inventory from the upper action limit to minimum level within the three-month period.

These formulas take into consideration lead times of manufacture and also you will note that action is taken gradually over a period of time. The reason for this is that our demand function is predicated on some prior period and any changes may or may not be significant. Also, we do not want to disrupt the shop with the shock of sudden changes in production. Thus, the time element in correcting production enables us to continuously re-evaluate demand and at the same time soften the impact of changing production schedules on the shop. Chart 4 shows graphically how the plan works. You will note that nothing is done about production for the first four periods even though demand is subject to some rather erratic swings. In the fifth period, the lower action limit is pierced. At this point, a change in the production rate is required, but since it takes two weeks to turn the production faucet on, inventory continues to drop for an additional week before turning around. I might say that each product is controlled in precisely this manner. In effect, we have some 250 separate plans. Also, this procedure was developed with pro-

gramming on an IBM 650 in mind.

Charts 5 and 6 show how the plan has really worked and use two real life examples. Looking at Chart 5, first, note the jagged, but continual decline in demand in the solid line. Second, note the erratic actual production curve, with production continuing until the first period in May. Third, note the wave-like upward sweep of the actual inventory which reached its peak in the May 1 period. Now, turn to the record as the plan, then under observation, would have performed. Actual demand, of course, is the same. But observe the dashed line labeled "Production (Based on Plan)"; it moves steadily downward from the November 1 period and would have cut off production in the February 3 period. Finally, inventories based on the plan would have peaked somewhat higher than in the actual case, but much earlier---in the December 1 period---and would have thereafter declined steadily, and by the June 4 period would have been in control.

Chart 6 portrays the actual versus the plan's production record for another one of the 250 products during the period of testing the plan. The dashed line shows the actual demand. The solid line shows the actual production schedule. Note its very volatile performance. Now, observe the way the plan would have dictated the production. These two examples were enough to satisfy the plant management that the techniques should be used. It has continued to give a good account of itself since last July, as demand has increased some 50 per cent.

In concluding this case, I would like to add that substantial inventory reductions have been realized along with greater production stability. Also, this plan has been far more sensitive in detecting and predicting demand changes than any other prior device, including empirical human judgment, available to the factory.

The subject of the second case might appear at first glance as not having too much to do with either Operations Research or Quality Control. This study concerns itself with the development of merit awards for the reduction in lost time accidents. Our Company has always been safety conscious and fully cognizant of the losses incurred both to the employee and to the Company from a lost time accident. In order to stimulate safety awareness we have, from time to time, offered prizes for going a specific period of time accident free. One of our divisions approached us and asked us to develop a plan which would be continuous in nature and in which prizes would be offered to the employees for good safety records. In essence, management wanted to know how many accident-free days should elapse before an award should be given and what the total costs of such a plan would be to the Company.

If the problem of accident spacing and their probability of occurrence is considered from a statistical viewpoint, it becomes fairly obvious that a model based upon the techniques of continuous sampling could go a long way toward explaining these factors. Chart 7, the first in this series, shows how the probabilities can be computed in this problem. You will, of course, recognize that this is a straight adaptation of the formulations in H. F. Dodge's "A Sampling Inspection Plan for Continuous Production" published in The Annals of Mathematical Statistics, Volume 14, No. 2, pp. 264-277, September, 1949.

Chart 8 shows the cost equation developed for evaluating the breakeven points between the costs of accidents and the awards. Before I get into this chart, let me give you some more background. First, management agreed that the breakeven point should occur at a 75 per cent improvement in the lost time accident rate. By this, we mean that any improvement of less than 75 per cent (except 0) should result in savings to the Company, in terms of the indirect and direct costs associated with an accident. Any improvement in excess of 75 per cent could result in a loss to the Company, i.e., management was willing to pay out more in award costs than the savings achieved through lost time accidents. Secondly, the plan described here is a three-part plan. If n consecutive days go by without an accident, a first prize is awarded. If 2n days are accident free, a better prize is awarded, and finally at 3n days, a grand prize is given. Any accident causes reversion to the starting point at whatever stage the group may then be. In using the equation shown, the probabilities of going n days and multiples thereof are evaluated (parts of overall equation in sub-brackets) and subtracted from the current cost. The .25 shown throughout represents a 75 per cent improvement. As you can see, the difference between the costs should equal 0.

Of course, there is only one problem with the above equation as it now stands. It cannot be solved mathematically. However, it is possible to arrive at an empirical or iterative solution. This is accomplished in the following manner. For each group at a factory location (and there are usually many departments competing), the appropriate p and q values are computed based on past history. A value of n is substituted initially in an attempt to satisfy the equation. Iteratively other values of n are used until one is found that gives a value of almost 0. This figure is then plotted against exposure or man-hours worked for each group. The X's on Chart 9 represent such solutions. A line is then fitted to the data and strata values are developed. The o's represent the values of n for each stratum. This can be better seen on Chart 10. In this particular case, then, we say that a group working 50,001 to 100,000 hours per month must go 150 days without a lost time accident to receive an award. On the other hand, a group working 400,001 to 500,000 hours per month need go only 50 days without a lost time accident.

Now, with regard to evaluating the program, Chart 11 shows the basic information required. Chart 12 shows how all of the formulas developed are used. You will note that awards are sometimes paid even though no improvement takes place. This is strictly a matter of probability. Also, in this case, at the 75 per cent improvement level, the Company has a saving of $1,865. We can attribute this saving to the stratifying process where a modified form of averaging takes place. Theoretically, this value should be 0.

I might add that this approach is being considered in a number of different areas of company activity. The exact formulations of the problems will vary with the desires of management. In one case the awards are continuous at the same value for each n days. In another case a grand award is given for each multiple of n greater than 3. In the case outlined above, the group reverts to the first award after achieving the grand award and recycles until a lost time accident occurs. Thus, we can see that the technique has great flexibility in conforming to management's desires. In one of our divisions, substantial savings

have been realized, and it can be proven conclusively that the improvement attained resulted primarily from the introduction of the plan. Also, in our more recent projects, hazard indices have been included with our exposure hours to make the award plan even more equitable as among competing groups.

Before I conclude, I would like to show some areas in which my shop has employed the tools of Operations Research rather successfully.

OBJECTIVES OF INVENTORY CONTROL AND PRODUCTION SCHEDULING PLAN

1. DETERMINE INVENTORY LEVELS REQUIRED BY "FREE MARKET" OPERATIONS
2. STABILIZE TOTAL PRODUCTION SCHEDULE BY ALLOWING THE INVENTORIES TO FLUCTUATE WITH MODERATE DEMAND VARIATIONS
3. MAINTAIN THE BEST POSSIBLE SERVICE TO THE CUSTOMER (MINIMIZE POSSIBILITY OF STOCK OUTS)
4. PROVIDE PROTECTION AGAINST OVERSTOCK CONDITIONS
5. DETERMINE WHEN AND WHAT QUANTITY TO ORDER
6. SCHEDULE PRODUCTION ON AN ECONOMIC BASIS
7. INDICATE NECESSITY FOR PRODUCTION REVISION WHEN THERE IS A SIGNIFICANT CHANGE IN DEMAND

DEMAND

INVENTORY CONTROL AND PRODUCTION SCHEDULING PLAN

GROUP	AVERAGE WEEKLY DEMAND	PRODUCTION CYCLE	CONTROL PERIOD	STANDARD DEVIATION FACTOR
1	1-7	6 MONTHS	MONTHLY	.90
2	8-25	3 "	"	.90
3	26-75	2 "	"	.90
4	76-150	MONTHLY*	"	.90
5	151-300	MONTHLY*	"	.50
6	301-500	MONTHLY*	"	.25
7	501-5,000	2 WEEKS	2 WEEKS	.15
8	OVER 5,000	WEEKLY	WEEKLY	.10

*PRODUCED ONCE A MONTH IN A SPECIFIED WEEK

INVENTORY CONTROL AND PRODUCTION SCHEDULING PLAN FORMULAE

INVENTORY CONTROL LIMITS

LOWER ACTION	$k_rD + Sk_rD$
MINIMUM	$k_rD + 2\ Sk_rD$
MAXIMUM	$k_rD + 4\ Sk_rD$
UPPER ACTION	$k_rD + 5\ Sk_rD$

WHERE:

D = AVERAGE WEEKLY DEMAND
S = STANDARD DEVIATION FACTOR
k_r = NUMBER OF WEEKS IN CONTROL PERIOD BASED ON NEW AVERAGE DEMAND

REVISION OF INPUT RATE

INVENTORY BELOW REQUIRED AMOUNT -

$$A_r = k_rD + \frac{k_r}{12}\left[(k_rD + 2\ Sk_rD) - (I + A_o - k_oD)\right]$$

ACTUAL OR ANTICIPATED OUT OF STOCK SITUATION

$$A_r = k_rD + \frac{k_r}{12}\left[k_rD + 2\ Sk_rD\right] - \left[I + A_o - k_oD\right]$$

INVENTORY EXCEEDS UPPER LIMITS (USED TO CORRECT FOR SIGNIFICANT DEMAND CHANGE)

$$A_r = k_rD - \frac{k_r}{12}\left[(I + A_o - k_oD) - (k_rD + 4\ Sk_rD)\right]$$

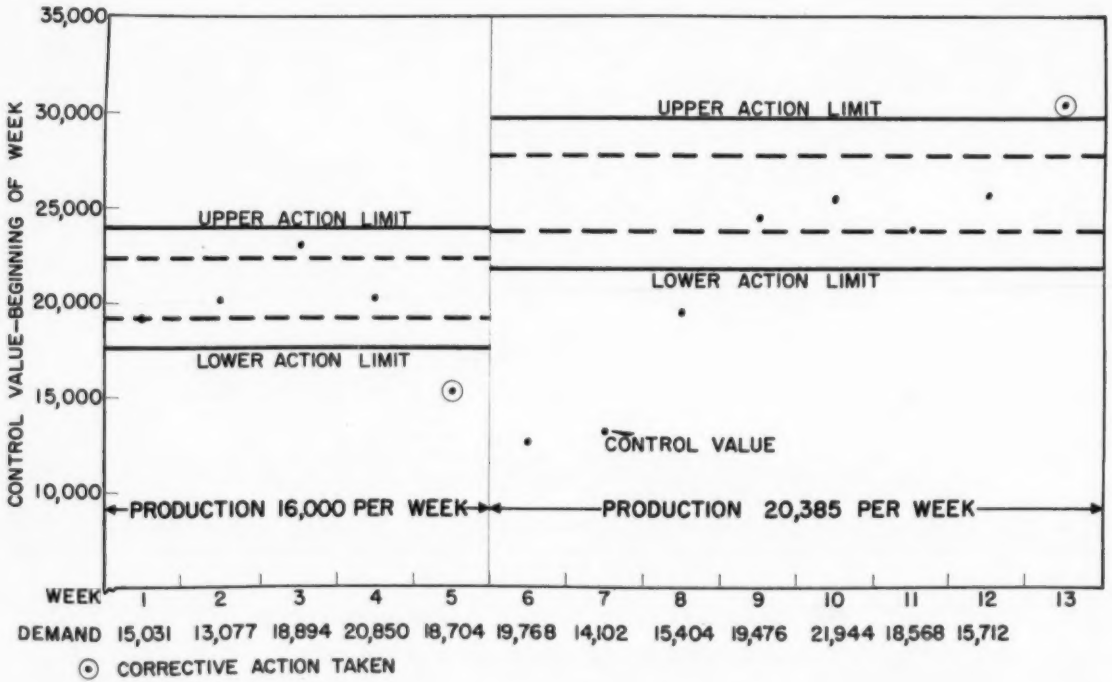
INVENTORY EXCEEDS UPPER LIMITS (USED TO CORRECT POTENTIALLY SERIOUS OVERSTOCK CONDITION)

$$A_r = k_rD - \frac{k_r}{4}\left[(I + A_o - k_oD) - (2\ k_rD + 6\ Sk_rD)\right] - \frac{k_r}{12}\left[k_rD + 2\ Sk_rD\right]$$

WHERE:

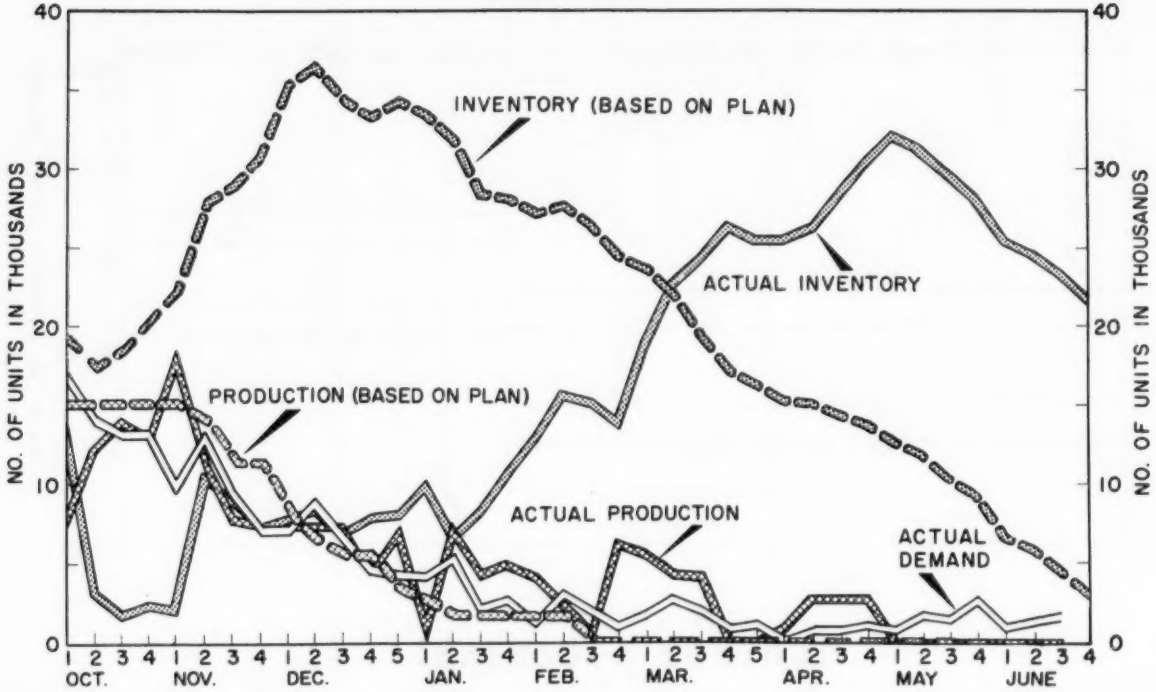
A_r = REVISED INPUT RATE
I = CONTROL VALUE
A_o = INPUT RATE FOR CURRENT PERIOD
k_o = NUMBER OF WEEKS IN CURRENT CONTROL PERIOD

INVENTORY CONTROL
PRODUCT A



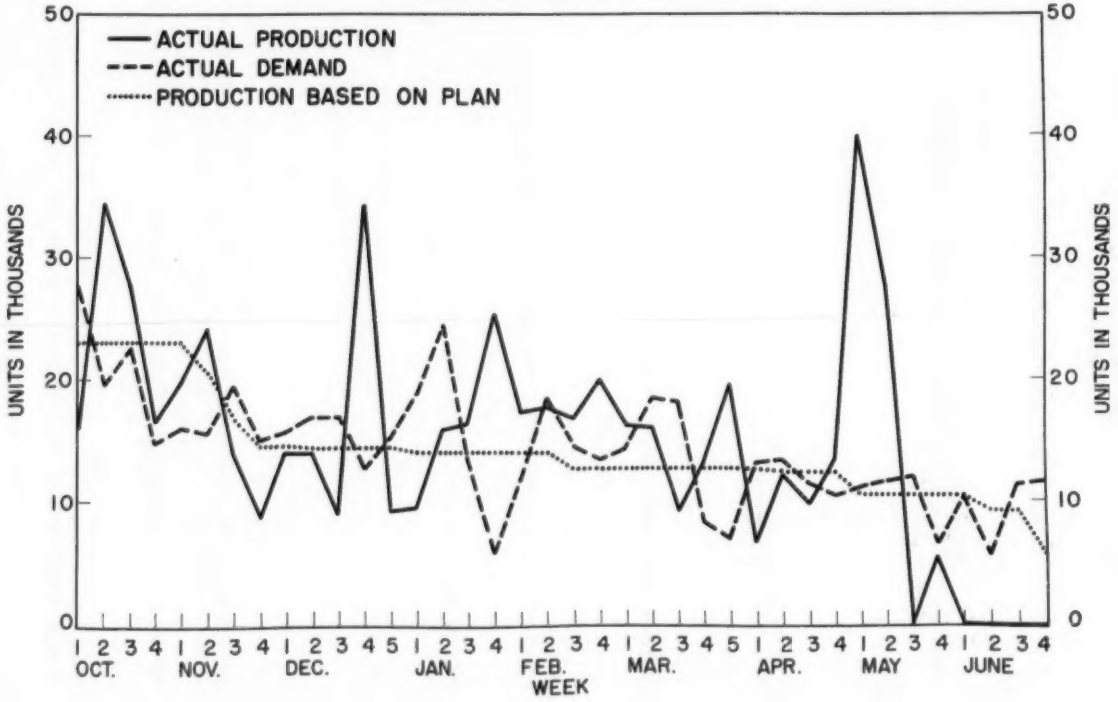
DEMAND, PRODUCTION, AND INVENTORY FOR 500DR-50 IVORY GENERAL PURPOSE TELEPHONE SET

ACTUAL VS. PLAN



500 DR - 3

INVENTORY PLAN VS. ACTUAL



DEVELOPMENT OF FORMULAE USED IN EVALUATION OF SAFETY AWARD PLAN

PROBABILITIES OF OCCURRENCE OF ACCIDENT SPACING

SEQUENCE	ACCIDENT SPACING	NUMBER OF ACCIDENT FREE DAYS PRIOR TO FINDING AN ACCIDENT	PROBABILITY OF OCCURRENCE	NUMBER OF TERM IN THE POWER SERIES
x	1	o	p	1st
ox	2	1	pq	2nd
oox	3	2	pq^2	3rd
ooox	4	3	pq^3	4th
.				
.				
.				
ooo...ox	n+1	n	pq^n	(n+1)st
.				
.				
.				

INFINITE POWER SERIES OF ABOVE PROBABILITIES

$$p + pq + pq^2 + pq^3 + \ldots\ldots$$

or

$$p\,(1 + q + q^2 + q^3 + \ldots\ldots$$

PROBABILITY OF FAILING TO FIND THE NEXT n DAYS FREE OF ACCIDENTS

$$P_1 = \sum pq^n = 1 - q^n$$

AVERAGE NUMBER OF DAYS IN A FAILURE SEQUENCE

$$h = \frac{p}{1-q^n} \frac{d}{dq} (1+q+q^2+q^3 + \ldots. + q^n)$$

$$h = \frac{1}{p(1-q^n)} \left[(1-q^n)\,(1+pn) \right]$$

WHEN $pn \longrightarrow o$

$$h = \frac{1}{p}$$

AVERAGE NUMBER OF FAILURE SEQUENCES BEFORE FINDING n DAYS FREE OF ACCIDENTS

$$G = Q\,(0+1P_1+2P_1{}^2+3\,P_1{}^3 + \ldots..)$$

$$G = Q_1\,P_1\,\frac{1}{(1-P_1)^2}$$

$$G = \frac{1-q^n}{q^n}$$

AVERAGE NUMBER OF DAYS PRIOR TO FINDING n DAYS FREE OF AN ACCIDENT

$$u = Gh + u$$

$$u = \frac{1-q^n}{pq^n}$$

DETERMINATION OF BOGEYS WITH BREAKEVEN POINT AT 75% IMPROVEMENT IN ACCIDENT RATE

$$\sum_{u=1}^{m} \left[pC_I - .25\ pC_I - \left(\frac{C_{A1} \times E \times .25\ p \times q^n}{(1-q^n)} \right) - \left(\frac{C_{A2} \times E \times .25\ p \times q^{2n}}{(1-q^{2n})} \right) - \left(\frac{C_{A3} \times E \times .25p \times q^{3n}}{(1-q^{3n})} \right) \right] = 0$$

WHERE:

u = UNITS PARTICIPATING IN THE CONTEST

p = ACCIDENT RATE PER DAY

C_I = TOTAL COST OF AN ACCIDENT

C_{A1} = VALUE OF FIRST AWARD

C_{A2} = VALUE OF SECOND AWARD

C_{A3} = VALUE OF THIRD AWARD

E = NUMBER OF EMPLOYEES IN THE UNIT

n = NUMBER OF ACCIDENT FREE DAYS REQUIRED TO WIN THE FIRST AWARD

q^n = THE PROBABILITY OF FINDING THE NEXT N DAYS WITHOUT AN ACCIDENT

DETERMINATION OF BOGEYS
BY MAN-HOURS WORKED STRATA

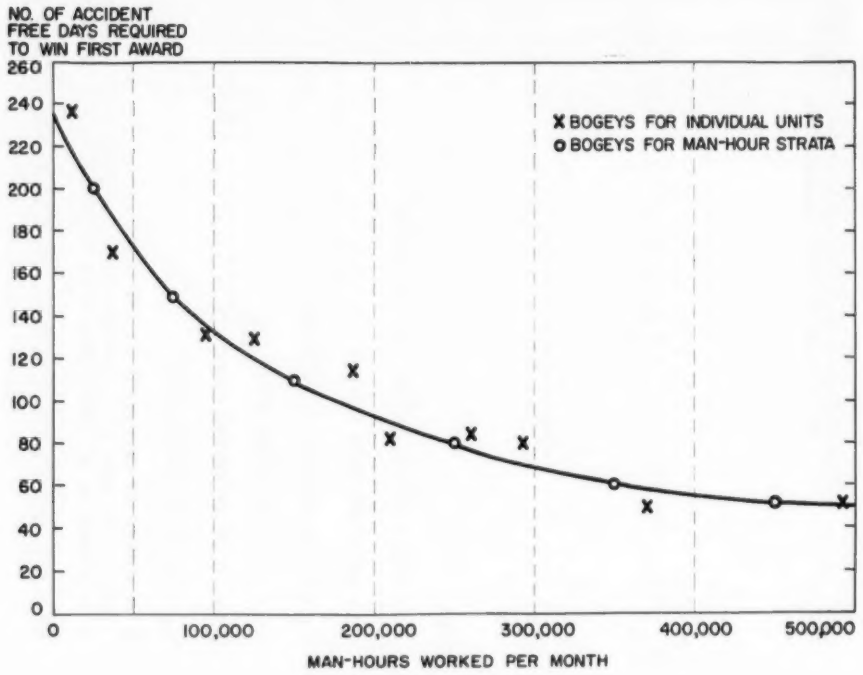


TABLE OF BOGEYS

MAN-HOURS WORKED PER MONTH	NUMBER OF ACCIDENT FREE DAYS REQUIRED TO WIN FIRST AWARD
0- 50,000	200
50,001-100,000	150
100,001-200,000	110
200,001-300,000	80
300,001-400,000	60
400,001-500,000	50

EVALUATION OF A UNIT IN THE SAFETY AWARD PLAN

NUMBER OF EMPLOYEES	1,500
MAN-HOURS WORKED PER MONTH	260,000
HISTORICAL ACCIDENT FREQUENCY	1.00 PER MONTH
COST OF AN ACCIDENT	$1,800.00
VALUE OF AWARDS	
FIRST	$2.00
SECOND	4.00
THIRD	6.00
NUMBER OF SUCCESSIVE ACCIDENT FREE DAYS REQUIRED TO WIN AN AWARD	
FIRST	80
SECOND	160
THIRD	240

FINAL EVALUATION OF A UNIT IN THE SAFETY AWARD PLAN

% IMPROVE-MENT	p	n	pn	q^{n*}	$1-q^n$	$\frac{1-q^n}{pq}$	NO. OF AWARDS UNTIL 3RD AWARD IS REACHED	COST PER AWARD	AWARD COSTS UNTIL 3RD AWARD IS REACHED	AWARD COST PER DAY	SAVINGS IN ACCIDENT COST PER DAY	NET DAILY SAVINGS	YEARLY SAVINGS
0	.0333	80	2.66	.0700	.9300	399.0	233.7	$3,000	$701,100				
		160	5.33	.0048	.9952	6,226.2	15.1	6,000	90,600				
		240	7.99	.0003	.9997	100,070.1	1.0	9,000	9,000				
									800,700	$ 7.95	0	($ 7.95)	($2,862)
25	.0250	80	2.00	.1353	.8647	255.6	54.0	3,000	162,000				
		160	4.00	.0183	.9817	2,145.8	6.4	6,000	38,400				
		240	6.00	.0025	.9975	15,960.0	1.0	9,000	9,000				
									209,400	13.12	$15.00	1.88	677
50	.0167	80	1.34	.2621	.7379	168.6	19.3	3,000	57,900				
		160	2.67	.0693	.9307	804.2	3.0	6,000	18,000				
		240	4.01	.0181	.9819	3,248.4	1.0	9,000	9,000				
									84,900	26.14	30.00	3.86	1,390
75	.0083	80	.66	.5175	.4825	112.3	4.50	3,000	13,500				
		160	1.33	.2647	.7353	334.7	1.3	6,000	7,800				
		240	1.99	.1367	.8633	760.9	1.0	9,000	9,000				
									30,300	39.82	45.00	5.18	1,865
100		80				80.0	1	3,000	3,000				
		160				160.0	1	6,000	6,000				
		240				240.0	1	9,000	9,000				
									18,000	75.00	60.00	(15.00)	(5,400)

*FROM POISSON TABLE

p = ACCIDENT RATE PER DAY
n = NUMBER OF ACCIDENT FREE DAYS REQUIRED TO WIN THE FIRST AWARD
q^n = THE PROBABILITY OF FINDING THE NEXT n DAYS WITHOUT AN ACCIDENT

SOME OPERATIONS RESEARCH STUDIES CONDUCTED AT THE WESTERN ELECTRIC COMPANY

DESCRIPTION	REMARKS
DEVELOP ECONOMIC MANUFACTURING QUANTITIES AND OPTIMUM INVENTORY LEVELS FOR PIECE PART SHOPS	ECONOMIC LOT SIZE PRODUCTION WAS INTRODUCED ON A WIDE VARIETY OF PIECE PARTS AND MACHINES.
DEVELOP A FACTOR WHICH WILL REFLECT THE COST OF EXTENDING TRANSPORTATION INTERVALS ON A DAILY BASIS	A FACTOR WAS DEVELOPED TO DETERMINE THE ADDED COST OF INCREASED SAFETY STOCK IN DISTRIBUTING HOUSES WHEN REDUCED TRANSPORTATION COSTS ARE ACHIEVED BY INCREASING THE TRANSPORTATION INTERVAL.
DETERMINE OPTIMUM DISTRIBUTION PATTERNS TO MINIMIZE LAID DOWN COSTS FOR PURCHASED ITEMS	OPTIMUM DISTRIBUTION PATTERNS WERE DEVELOPED BY LINEAR PROGRAMMING ANALYSIS WHICH YIELDED THE LOWEST LAID DOWN COSTS.
LOCATION OF NEW PLANT	DETERMINE OPTIMUM LOCATION FOR NEW PLANT TO MINIMIZE TRANSPORTATION COSTS CONSIDERING RAW MATERIAL PROCUREMENT AND DISTRIBUTION OF FINISHED PRODUCT.
LINEAR PROGRAMMING ANALYSIS OF EXCHANGE CABLE PRODUCT MIX	SEVERAL ALTERNATE COMBINATIONS OF CABLE TO BE PRODUCED AT THE ABOVE PLANT WERE ANALYZED TO DETERMINE MINIMUM TRANSPORTATION COSTS.
DEVELOP SAMPLING PLANS FOR AUDITING	SCIENTIFIC SAMPLING EFFECTED A REDUCTION OF 55% TO 90% IN AUDITING EFFORT WHILE MAINTAINING QUALITY OF AUDIT.
DETERMINE MAXIMUM ERROR THAT CAN BE EXPECTED USING AN AVERAGE SPOOL WEIGHT IN DETERMINING IN PROCESS INVENTORIES	THIS STUDY WAS MADE TO DEVELOP A MEANS FOR ADJUSTING IN PROCESS INVENTORIES OF WIRE FOR WAGE INCENTIVE PURPOSES TO COMPENSATE FOR VARIANCE IN SPOOL WEIGHTS.
FORECAST OF BUSINESS ACTIVITY FOR PROGRAM ANALYSIS	FORECASTS WERE DEVELOPED BASED ON VARIOUS NATIONAL ECONOMIC SERIES.
DEVELOP PLAN TO DETERMINE MAINTENANCE POLICY FOR INCLINED ENAMEL MACHINES	ESTABLISH OPTIMUM MAINTENANCE AND CLEANING INTERVALS BASED ON ECONOMICS AND CONSIDERING PRODUCTION LEVEL, TYPE AND QUALITY OF PRODUCT.

STATISTICAL PROCEDURES APPLIED TO PROCESSED FRUITS AND VEGETABLES

Richard P. Bartlett, Jr.
Agricultural Marketing Service
U. S. Department of Agriculture

The Fresh and Processed Products Standardization and Inspection Branches of the Fruit and Vegetable Division, Agricultural Marketing Service are engaged in two major fields of activity. One is the development of U. S. Grade Standards for fresh fruit and vegetable and for processed fruit and vegetable products and revising those standards when the need arises. The other is making inspection service available to industry and Federal Government and State procurement agencies. In performing these activities destructive sampling of product is unavoidable on the processed products and in many instances on the fresh products. This suggests that "variables" type of sampling is desirable in order to keep the sample size small. However, many of the statistical procedures used for inspection of fruit and vegetable products such as those presented herein are confined to sampling by "attributes". The reasons for use of "attributes" are as follows:

1. "Variables" plans are very difficult to administer nationally because statistical training of inspection personnel would be essential. This objection is gradually disappearing due to in-service training and the advent of more "variable" acceptance sampling procedures (6).

2. Many factors tested are not always measurable in the variables sense, such as flavor, color, and maturity. This limitation is being reduced due to the development of objective instruments by the various research groups.

3. The present scoring system as applied to processed food does not follow the normal distribution theory, necessitating special theory application of "variables" sampling which at present is beyond the scope of practical consideration. As an illustration, consider the quality inspection of a processed product in which a lot or population is assigned U. S. Grade A, B, or C, or denoted as "Substandard" or "Grade not Certified." A sample unit is assigned a grade based on score points such as 90-100 for A, 80-89 for B, and 70-79 for C; but may be assigned a lower grade than indicated by the score under certain limiting conditions which are prescribed for a specific product. /For example (2)/ It is possible with these limiting conditions for sample units scoring 90-100 points to be classified anywhere from U. S. Grade A to "Substandard", thus the score points do not always reflect the acceptability of a sample unit. As a result, normal theory in relation to the Grade is not applicable to such variables.

   From the theoretical viewpoint, this objection can be remedied by changes in scoring methods but the scoring methods cannot easily be changed. The time, manpower and educational program required, including the obtaining of industry views, for such

a change is insurmountable at present.

Therefore, confronted with these facts it seemed most practical to define each sample unit as defective or non-defective for a particular grade and adopt for the present what was already in use, attribute acceptance sampling. The sampling terms "defect" and "defective" were replaced by "deviation" and "deviant" respectively to avoid conflict with other previously established usage of the words "defect" and "defective" in the U. S. Grade Standards.

For the purpose of quality inspection of processed fruits and vegetables to which this discussion applies, a "deviant" was specifically defined as a sample unit that falls into the next grade below the indicated grade but does not score more than 4 points below the minimum total score for the indicated grade. For example, if the indicated grade was A, and the score for A was 90-100, a sample unit scoring 91 points but assigned Grade B due to a limiting rule would be classed as a deviant. Likewise, a sample unit scoring 86 to 89 points would be a deviant when the Grade A requirement is considered.

The Acceptable Quality Level (AQL) for such deviants was established at 6.0% with the producer's risk (α) equal to 5.0% or less. Single-sample plans were developed to satisfy the AQL and α, beginning with a sample size of 3. Since the plans had to be determined from a fixed AQL, α and n (sample size), the lot tolerance percent defective (p_t) for a fixed consumer's risk, β = 10%, varied with sample size. The larger the sample size the smaller p_t. This is shown in figure 1, which contains the operating characteristic (OC) curves for the single-sample plans so developed and are those in use today.

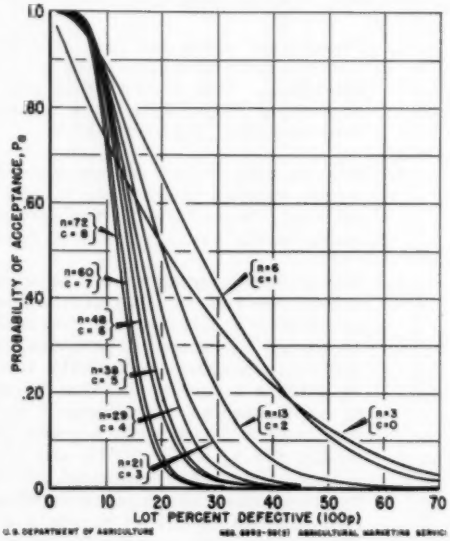


Figure 1. - Operating characteristic curves for quality single-sampling plans - AQL = 6.0 percent.

The different sample sizes are for use with different lot sizes. The larger the lot size and/or container size for the same product the larger the required sample size, thus decreasing the p_t as the value of the lot increases. Table 1 illustrates the allocation of sample size with respect to lot and container size for certain

canned fruits and vegetables. The sample sizes are required minimums only; any time it is deemed desirable a larger sample size corresponding to one listed in the table may be selected as well as sample sizes provided elsewhere (1) up to n = 400.

TABLE 1

Canned or similarly processed fruits, vegetables, fishery products, and products thereof containing units of such size and character as to be readily separable

Container size group	Lot size (number of containers)								
GROUP 1 Any type of container of less volume than that of a No. 300 size can (300 x 407).	3,600 or less	3,601-14,400	14,401-48,000	48,001-96,000	96,001-156,000	156,001-228,000	228,001-300,000	300,001-420,000	Over 420,000
GROUP 2 Any type of container of a volume equal to or exceeding that of a No. 300 size can, but not exceeding that of a No. 3 cylinder size can (404 x 700).	2,400 or less	2,401-12,000	12,001-24,000	24,001-48,000	48,001- 72,000	72,001-108,000	108,001-168,000	168,001-240,000	Over 240,000
GROUP 3 Any type of container of a volume exceeding that of a No. 3 cylinder size can, but not exceeding that of a No. 12 size can (603 x 812).	1,200 or less	1,201- 7,200	7,201-15,000	15,001-24,000	24,001- 36,000	36,001- 60,000	60,001- 84,000	84,001-120,000	Over 120,000
GROUP 4 Any type of container of a volume exceeding that of a No. 12 size can, but not exceeding that of a 5-gallon container.	200 or less	201-800	801-1,600	1,601-2,400	2,401-3,600	3,601-8,000	8,001-16,000	16,001-28,000	Over 28,000
GROUP 5 Any type of container of a volume exceeding that of a 5-gallon container.	25 or less	26-80	81-200	201-400	401-800	801-1,200	1,201-2,000	2,001-3,200	Over 3,200.
	Single sampling plans [1]								
Sample size (number of sample units). [2]	3	6	13	21	29	38	48	60	72
Acceptance number	0	1	2	3	4	5	6	7	8

[1] For extension of the single sample sizes beyond 72 sample units, refer to table VI of this section; for multiple sampling plans comparable to the various single sampling plans refer to table VII of this section.

[2] The sample units for the various container size groups are as follows: Groups 1, 2, and 3—1 container and its entire contents. Groups 4 and 5—approximately 2 pounds of product. When determined by the inspector that a 2-pound sample unit is inadequate, a larger sample unit may be substituted.

The 6.0% AQL used for these plans is well suited to the plans in Military Standard 105A for an AQL of 6.5%, but the sample sizes provided therein together with the applicable lot sizes were not very easily adaptable to the pre-established rates and methods of sampling used by the Department of Agriculture for processed fruits and vegetables. The "jumps" from one sample size to the next larger were held as low as possible by USDA for the required AQL and α due to the destructive sampling. A smaller p_t than those of Military Standard 105A was provided for sample sizes exceeding 25, while the p_t for sample sizes equal to or less than 25 is larger for USDA due to the necessity of keeping the "jumps" small. A further reason for departing from Military Standard 105A was that for an AQL of 6.5%, increasing the sample size as provided by Mil. Std. 105A sometimes resulted in increased chances of accepting lower quality product as evidenced by the change from n = 25, c = 3 to n = 35, c = 5; or n = 10, c = 1 to n = 15, c = 2. This is illustrated by the OC Curves of figure 2.

To further reduce the sample sizes, multiple sampling plans were developed to fit the OC Curves for the single sampling plans of figure 1, with the exception of n = 3, c = o. The reasoning behind these plans was to begin the first sample of the sequence as small as possible but large enough for acceptance and rejection to be permitted. Both acceptance and rejection were necessary because if the lot in question was, for example, rejected for Grade A on the first sample size of the multiple plan, it could very likely be accepted at that point for Grade B or lower depending upon the sample results. A second requirement was to truncate each plan as soon as possible so that the total sample size for multiple would not greatly exceed the corresponding single sample size. This was desirable in order to enable use of a multiple plan when the sample was drawn several miles from the place of inspection, requiring selection of the total sample during the initial sampling. The producer would readily discourage an inspector carrying away, say, fifty percent more product than needed for single-sampling in order to provide for using the multiple plan.

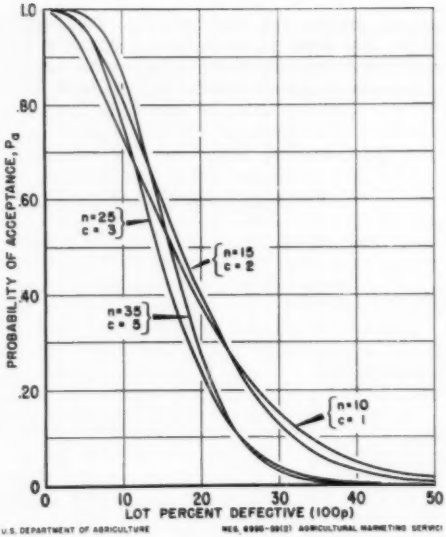


Figure 2. - Operating characteristic curves for some single sampling plans in Military Standard 105A - AQL = 6.5 percent.

Table 2 lists the multiple plans discussed above and their corresponding single sampling plans. These plans are, for the reasons given above, decidedly different from those of Mil. Std. 105A, an example of which is shown in figure 3.

TABLE 2

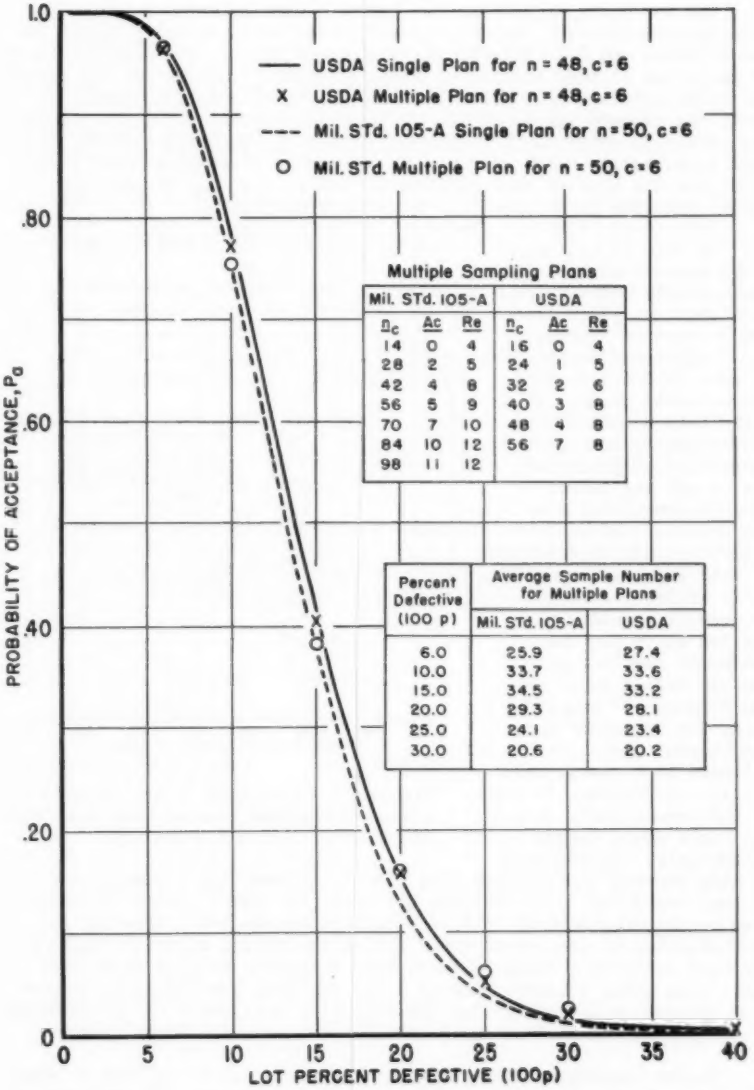
Multiple Sampling Plans[1, 2]

Multiple sampling plans comparable to the indicated single sampling plans

INDICATED SINGLE SAMPLING PLAN	Single sample size, n	6	13	21	29	38	48	60	72
	Acceptance numbers, c	1	2	3	4	5	6	7	8
Cumulative sample sizes, n_c, and acceptance numbers, c, and rejection numbers, r, for multiple sampling.		n_c c r 4 0 2 6 0 2 8 1 2	n_c c r 8 0 3 10 0 3 12 1 3 14 2 3	n_c c r 10 0 3 14 1 4 18 1 4 22 2 5 26 4 5	n_c c r 12 0 4 16 0 4 20 1 5 24 2 5 28 3 6 32 3 6 36 5 6	n_c c r 14 0 4 20 0 5 26 1 6 32 2 6 38 3 7 44 6 7	n_c c r 16 0 4 24 1 5 32 2 6 40 3 8 48 4 8 56 7 8	n_c c r 18 0 5 28 1 6 38 2 7 48 3 8 58 4 9 68 6 9	n_c c r 22 0 5 32 1 7 42 2 8 52 3 9 62 5 10 72 6 10 82 9 10

[1] The above multiple sampling plans may be used in lieu of the single sampling plans listed at the heading of each column.

[2] The last number in each n_c column is the total number of sample units that may be required for multiple sampling. However, it is possible to complete inspection with less sample units if the number of deviants (or deviations when so specified) does not exceed the acceptance number, or equals or exceeds the rejection number for the appropriate sample size.



Multiple Sampling Plans

Mil. STd. 105-A			USDA		
n_c	Ac	Re	n_c	Ac	Re
14	0	4	16	0	4
28	2	5	24	1	5
42	4	8	32	2	6
56	5	9	40	3	8
70	7	10	48	4	8
84	10	12	56	7	8
98	11	12			

Percent Defective (100 p)	Average Sample Number for Multiple Plans	
	Mil. STd. 105-A	USDA
6.0	25.9	27.4
10.0	33.7	33.6
15.0	34.5	33.2
20.0	29.3	28.1
25.0	24.1	23.4
30.0	20.6	20.2

U.S. DEPARTMENT OF AGRICULTURE NEG. 6997-59(2) AGRICULTURAL MARKETING SERVICE

Figure 3. - Comparison of characteristic curves for the indicated U. S. Department of Agriculture and Military Standard 105A sampling plans.

The two plans illustrated in figure 3 were selected because they were the closest with respect to sample size and acceptance number for the 6 (USDA) and 6.5 (Mil.Std. 105A) percent AQL's. The discrimination of the plans is very similar as shown by their OC curves and the curves for their corresponding multiple plans show good agreement also. In fact, the average sample number for each are fairly close. Therefore, the remaining criteria for selection of one of the plans are the initial sample size, total sample size and number of samples required by each plus the size of each additional sample. For our purpose the (USDA) multiple plans were felt to be more applicable.

In view of the high risks associated with the small sample sizes, particularly n = 13 or less, the concept of "reduced and "tightened" inspection as used in Mil.Std. 105A is now under study. It is hoped that a workable procedure can be adapted to inspection of processed fruits and vegetables thereby benefiting from the decreased risks as illustrated in figure 4. In this figure, the middle curve is for the normal or present sampling plan; the top curve for the "reduced" sampling plan; and the bottom curve for the "tightened" sampling plan. The "reduced" and "tightened" plans are intended to be used, under certain conditions, in lieu of the normal plan, thus the risks would change accordingly. In effect, if this concept were applied, the actual OC Curve for a sample size of 6, say, would look more like the top curve in figure 4 from 0% defective to the neighborhood of 6.0% defective in the lot. Whereas, the probabilities of acceptance would follow the bottom curve of figure 4 for lots containing slightly above 6.0% defective to 100% defective. Thus, when using a sample size of 6 and applying the tightened inspection concept, a lot containing 30% defective would have .11 probability of acceptance rather than .42 probability as with the normal plan.

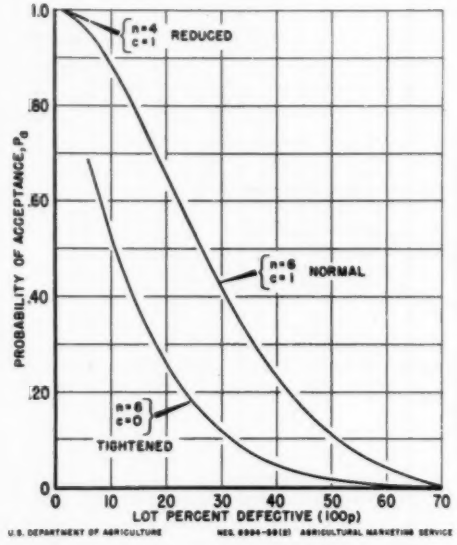


Figure 4. - Operating characteristic curves for normal, reduced and tightened inspection - AQL = 6.0 percent.

We are optimistic that statistical methods can be applied in many other ways in the inspection of fruits and vegetables. We realize, of course, that before such changes can be made a part of the everyday inspection practices, the statistical methods must be adapted to fit the economical and practical objectives of the inspection program.

REFERENCES

1. Agricultural Marketing Service, U. S. Department of Agriculture. Regulations governing inspection and certification of processed fruits and vegetables and related products. Effective July 1, 1958.

2. Agricultural Marketing Service, U. S. Department of Agriculture. United States Standards for grades of canned sauerkraut. Effective May 10, 1957.

3. Freeman, H. A. et al. Sampling Inspection, 1948. Mc-Graw-Hill Book Company, Inc., New York and London.

4. Grant, E. L. Statistical Quality Control. 1946. McGraw-Hill Book Company, Inc., New York and London.

5. Munitions Board Standards Agency, Department of Defense. Military Standard. Sampling Procedures and Tables for Inspection by Attributes (Military Standard 105A). Washington: Government Printing Office, 1950.

6. Supply and Logistics, Office of the Assistant Secretary of Defense. Military Standard. Sampling Procedures and Tables for Inspection by Variables for Percent Defective (Military Standard 414). Washington: Government Printing Office, 1957.

PRODUCTION VIBRATION PROCESSING

W. C. Hanna and W. H. Brown
Quality Engineering
Nortronics, A Division of Northrop Corporation

This paper discusses the vibration test program of the guidance system used in the SM-62 "Snark" intercontinental missile. It is a rocket launched, air-breathing missile guided by a stellar supervised inertial guidance system.

The guidance system is composed of approximately thirty "black boxes", each a major functional unit containing hundreds of components. It goes without saying, that for any reasonable probability of a successful mission, we must have a very high reliability of each of the components and their major unit assemblies. While functional and visual tests are made at many assembly levels on every system, we must also somehow provide coverage of the adverse environment in which the system will live.

Vibration tests were started more than two years ago on one of the prototype systems. A large number of design deficiencies were immediately uncovered. These were largely mechanical problems with component supporting structure. Insufficient attention had been given to designing for rigidity, and resulted in low frequency resonances of large amplitude. This caused electrical malfunctions as well as mechanical failures of components and structure. Design fixes were made on this particular set of specimens until they were proven capable of withstanding the environment.

The prototype systems were assembled, functionally tested, and environmentally tested by the Engineering Division, with the primary purpose of proving design adequacy.

As the first production systems were built, a great many problems developed in the course of assembly and test. As the engineers were called in to assist in their solution, they observed many small details not to their liking.

It must be recognized that the production version is a different breed. It is made from formal drawings, passes through many departments with different personnel. Fabrication and assembly techniques are different, many parts have been redesigned for each of production and/or environmental fixes, fasteners, components and wire routing are different.

It was decided that Engineering should review completed systems. Accordingly, each Responsible Engineer studied and functionally tested the production "black box" of his design. Literally, hundreds of discrepancies were found. A Quality Review Committee was then formed, with representatives of Engineering, Manufacturing, and Quality Control. Each of the discrepancies was classified as to cause and corrective action was initiated in the responsible departments.

This procedure made a vast improvement in the quality of the system, but visual inspection and functional testing are not enough; environment-

al tests must also be made. Certain classes of defects cannot be discovered any other way.

An Environmental Review program was inaugurated on a trial basis whereby approximately one-third of the completed systems were subjected to vibration testing.

On the first system tested there were 65 malfunctions of which 35 were critical and could have aborted a missile flight!

This same system had been very thoroughly checked for discrepancies by the Quality Review Committee and all discrepancies considered critical to the running of vibration tests were corrected prior to the test. Of the 35 critical malfunctions; 10 were component failures, 13 were caused by workmanship, 12 by inadequate design.

In all, 7 complete systems were tested. The number of design failures per system showed a steady decrease to zero, except for some well known repeating failures whose fixes were not scheduled until later systems. Likewise, the number of workmanship failures showed a steady decrease from 13 per system to 1 per system. However, the number of component failures varied between 5 and 13 per system.

These trends have considerable significance. As might be expected, failures caused by inadequate design can eventually be virtually eliminated. Failures caused by faulty workmanship can be reduced to the vanishing point by training programs, and a follow-up on every flaw found. Constant surveillance of the product, however, is necessary to keep it that way. The number of component failures per system did not decrease. They appeared only under vibration and could not be weeded out by the many functional tests made during assembly and the Quality Review of the completed system.

Now that our design problems are pretty well licked, our problem is to maintain workmanship quality at a very high level and to identify marginal components.

Accordingly, we are starting two new programs to control the quality of the product. The first is a Receiving Environmental Test of components composed of two phases:

(a) A non-destructive test primarily to check on vendor workmanship. The test will be to the purchase specification but attenuated in scope and severity.

(b) A destructive test to prove the components are meeting the full value of their purchase specification.

Both of these will be done by lot sampling.

The second new program is Production Line Vibration Processing. We call it a process rather than a test because it will be conducted by Manufacturing rather than Engineering and will be handled in a formal routine manner as any other manufacturing process. It will provide a close surveillance of workmanship quality and component quality by detecting flaws not previously found by the many inspections and functional

tests.

It is realized that there is considerable controversy over the merits of production vibration testing. We are, therefore, vibrating all of the "black boxes" but with only half of them done on alternate systems. By this method we can make comparisons of service malfunctions between those units that were vibration processed and those that were not.

In the matter of vibration criteria there is the controversial problem of how to test without damaging the specimen but still detect flaws in it. We have done several things to reduce vibration time to avoid fatiguing the specimen:

(a) Only one axis is tested. It has been found that, because of complexity of the equipment, its response is similar regardless of the input axis.

(b) Resonance testing is not used. Malfunction can usually be detected when sweeping through a resonant frequency. Putting additional time on the specimen at the worst frequency serves no useful purpose here. Resonant problems generally involve supporting structure and have been well covered in early tests.

(c) The production systems tested to date have been exposed to four identical 15 minute sweeps. It was found that the malfunction rate definitely decreased with the number of sweeps. Most occurred in the first sweep, some in the second, and almost none in the third and fourth. Our new program will have only two 15 minute sweeps. The actual flight then is equivalent to the third and fourth sweeps where we know the failure rate is low.

Selecting an acceleration level is something of a problem. It must be high enough to detect flaws in the equipment, but not so high as to damage the equipment or to create abnormal malfunctions. At one point in the program it was thought that a high acceleration level would be the best and quickest way to shake down a specimen. However, it became clear that the design limitations were being exceeded; the failure rate was very high and failures of a design nature were occurring that were not typical of previous experience.

We are, therefore, using the level required by the missile qualification specifications. This is somewhat higher than the actual flight environment, but the main source of safety in avoiding permanent damage to the equipment is one of time rather than of acceleration. The capability of a good system to withstand lengthy vibration tests was demonstrated in another program, that of full Qualification Tests to all environments. In that program, failures were reworked as they occurred and the test continued to completion. The cycle was then begun again and carried to the same point as the installation of each rework; thus each change was tested a full cycle and its time capability was demonstrated.

Figure I illustrates the acceleration level we are using. Each of these sweeps, Curve A and B, required 7 minutes and is done twice for a

total time of 28 minutes. The frequency is swept logarithmically with time.

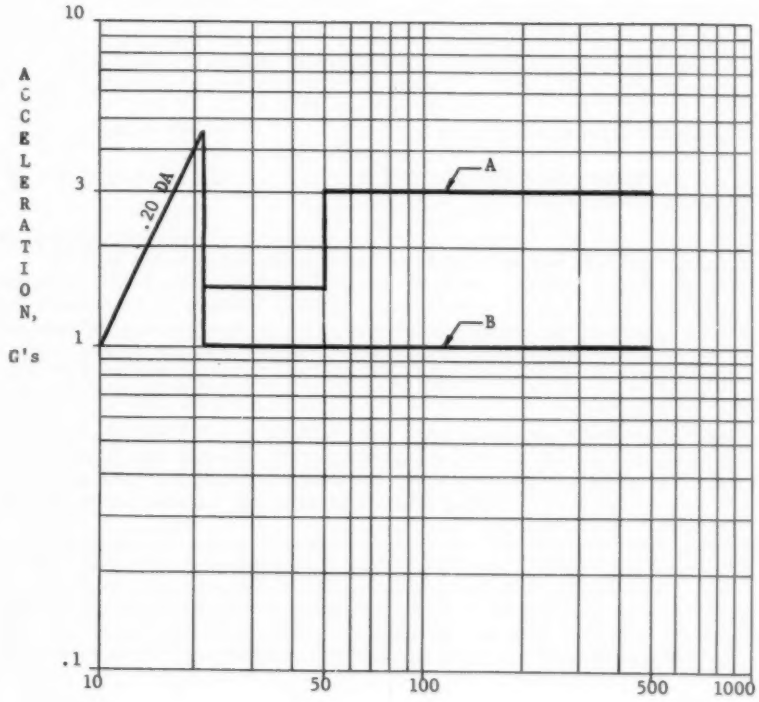
The system equipment is mounted in a large rack which is isolated from the missile by non-linear rubber isolators with a natural frequency of 16 cps. Since the specimens are more easily tested individually, the presence of the isolators in the next assembly is reflected in the shape of the test curve. The constant displacement line between 10 and 21 cps allows for resonance of the rack isolators, and because they are non-linear, there is a peak response and sharp cut-off at 21 cps. Since the specimens have no resonances below 50 cps they are little affected by this portion of the test. However, loose or missing fasteners are readily detected and any foreign material is shaken out. The remainder of the curve allows for 20% transmissibility through the isolators. The higher portion of Curve A simulates missile launch where there is a high vibration level for a short time from the booster rockets.

As mentioned earlier one of the major problems is adequate control of components. One of the causes of this is the specification problem. Components are purchased under a variety of specifications with different environmental conditions. These range from MIL specifications which don't agree; to vendor's catalogs which quote some parts of several MIL specifications, or their own; to commercial parts with no environmental specification at all.

Then, there is the specification for our end product which we must meet. This applies several assembly levels up from the component; which is a particularly important point in the case of vibration because the dynamic response changes as soon as two parts are put together.

We are endeavoring to write environmental specifications for components that will assure satisfactory operation when assembled into our end product. In the meantime, the Production Line Vibration Process is expected to close this large gap by checking all components as well as workmanship of the assembly.

FIGURE I


10
3
1
.1
ACCELERATION, G's
.20 DA
A
B
10
50
100
500
1000
FREQUENCY, cps

QUALITY CONTROL ENGINEERING IN NEW DESIGN

William J. Masser
General Electric Company

New Design offers an opportunity for the Quality Control Engineer to do solid, long range planning in prevention of quality problems. By devoting careful attention to details early in the product development program, much trouble-shooting is made unnecessary during production.

The preventive effort of the quality control engineer for new designs is directed at:

1. The Product Design)		(A. Pre-Production Quality Evaluation
)	for	(
2. The Process Design)		(B. Product and Process Quality Planning

Pre-Production Quality Evaluation

A close team relationship must exist between the quality control engineer, the product design engineer, and the process design engineer for evaluation of product quality prior to the start of manufacture. The quality control engineer can contribute toward the analysis of the product design for conditions that might cause quality problems. He does this on the basis of his experience and knowledge obtained from previous similar designs.

The quality control engineer can also assist in identifying the environments and end-use conditions against which the product should be evaluated. Analysis of field results and customer complaints with earlier models equips him with this knowledge.

Although the design engineer usually has the primary responsibility for evaluation of prototype quality, the quality control engineer's testing experience may assist in providing helpful suggestions. The quality control engineer also benefits from a close working relationship during prototype testing. It provides him with valuable background experience toward planning in-process testing of the manufactured product.

Before the product design engineer can complete his product design he finds it necessary to work with the process design engineer in deciding what manufacturing process will be used to make the part. The selection of the process may depend on the many factors such as design requirements, manufacturing costs, availability of facilities, process capabilities, and the like. A compatible relationship must be established between product requirement and process capability. The quality control engineer can contribute toward the team effort in this activity by analyzing tolerances. He may find it possible to partition tolerances so that by tightening the tolerance on an easy-to-make part, the tolerance could be "opened up" on a difficult part. He can also conduct process capability analyses and special process studies as required.

During the review of drawings and specifications, the quality control engineer should be alert for any quality requirements that might be misinterpreted or misunderstood by vendors, foremen, operators, inspectors or engineers. Points where clarification is needed are brought to the attention of the design engineers for correction.

Product and Process Quality Planning

Another important task that also requires a team effort is Product and Process Quality Planning. The quality control engineer has the primary responsibility for getting the planning done and documented. The product design engineer can give valuable assistance in classifying quality characteristics as to their relative importance with respect to quality of design - including performance and reliability. The product design engineer may also specify certain tests to assure that the product and process design is producing the desired end-result. The quality control engineer is responsible for planning the detailed test procedures and the test equipment.

Quite obviously, quality planning is dependent upon the manufacturing process since much of the quality control effort is directed at controlling of processes. This calls for a close working relationship between the quality control engineer and the process design engineer. This relationship becomes increasingly important with increased mechanization and automation of processes. Under these circumstances, quality measurements and controls must be closely integrated with the manufacturing process.

When both the product design and process design have been completed it is possible to design the quality control system. The quality control engineer determines what quality characteristics should be measured and when these measurements should be made in the process flow. He also plans the quality information flow. This should include instantaneous measurements fed back to the operator so he can tell the quality he is producing at the time he is producing it. It should also include a measure of composite quality of the final product - a quality index - for managerial control of the business system.

Other details covered by quality planning include jig and fixture control, calibration of quality measuring equipment, qualification of personnel, training requirements and the like.

The job of Product and Process Quality Planning is not complete until the detailed procedures have been documented in a clear, concise manner, and provisions made to revise the procedures as found necessary.

The quality control engineer can make a significant contribution to the industrial community in particular, and to the public in general, by preventing the occurrence of quality problems through careful analysis and planning during the design stage.

NAVY PROGRESS IN QUALITY CONTROL FOR 1958

Captain George L. Heap, USN
Assistant Chief of Naval Material (Field Services)
Office of Naval Material
Department of the Navy

My remarks today are directed to the progress NAVY has achieved during 1958 in the "production inspection" aspects of QUALITY CONTROL. Vigorous application of Value Engineering Programs, wide circulation of quality data, and expedited handling of feed-back information from inspecting and user activities have produced gratifying improvements in design and in product evaluation systems. These quality areas are too broad to be covered adequately within the time allotted for this presentation.

Navy's most significant progress during 1958 in the area of production inspection derives from a single major change in stated policy: a shift in emphasis from inspection, in the sense of sorting the good from the bad, to quality control, in the sense of effective control of quality built into the production processes. This restatement of policy is no "bolt-from-the-blue" as far as Navy's basic inspection ideas are concerned. For 165 years, Navy has never deviated from the concept that we can be assured of ultimate end item functional performance of complex products under industrial contracts only by continuous governmental surveillance at production source throughout the contractor's production processes. The necessity for a clarifying restatement of governing policy has become urgent during this current period of accelerated technological advancement in weapons and weapons systems. All old methods, time-honored though they may be, are not suited to complicated modern equipment, such as missiles and nuclear propulsion systems. Most missiles cannot be given a trial run or a flight test, our former procedure with ships and aircraft, without destruction. Nor can we afford the extravagance of enough extra shots to compensate for the percentage we know will miss the mark, as we did with conventional, less expensive ammunition. At today's prices, we must be assured, by continuous production surveillance, that producer's methods and procedures will deliver to us weapons manufactured to our specifications and capable of performing functions for which they are intended.

Considering the technical complexities of the problem as well as the normal, healthy differences in the opinions of "experts," I was not surprised that Navy's revised inspection policy was received in the field with widely varying reactions. Opinions ranged from: "now we will get nothing but junk" to "now we will get what we are paying for without doing the contractor's job for him." My personal views lean toward a more moderate reaction: "we have been doing that for years - we never emphasized inspection production of the good producers to the same extent as that utilized for the marginal producers. What was done by this directive was to make policy of a practice that made good sense. Despite this lack of uniformly favorable field acceptance, restatement of policy did accomplish one highly important operating objective. It removed any doubt that Navy

expects its suppliers to assume contractual responsibility to establish procedures assuring production of material meeting specified quality, and to generate and make available to Government representatives for use in determining product acceptability quality control data adequate in detail for this purpose.

Implementing major technical procedural changes such as this takes time and sustained guidance. You're bucking the normal human inertia which resists any change. By the close of 1958, the ever-increasing momentum of full acceptance of the new policy's intent had achieved substantial progress, and its beneficial effects were in evidence. I give our field people full credit for these accomplishments. They reflect effective thought and effort to surmount factual operating problems encountered in at least four different categories, depending upon factors involved in individual contracts under administration. These may be listed as follows:

1. Existing contracts under which production was underway and satisfactory in all respects;
2. Existing contracts under which production was underway but not satisfactory in all respects;
3. New contracts not incorporating the requirement that contractor employ quality control procedures conforming to revised policy; and
4. New contracts stipulating the revised quality control procedures, with or without revised specifications.

The first area covered existing contracts for which production was underway and satisfactory in all respects. If the contractor was receptive to installing the new quality control procedures without an increase in the cost of his contract, and if there was enough production remaining to make it worthwhile, steps were taken to accomplish this. If he was not agreeable, the contract was permitted to run its normal course to completion. Considerable progress was made by our field offices in connection with this area of contracts. The offices have shown good judgment and initiative in determining just how surveillance inspection was to be accomplished on a contract by contract basis. Field efforts along this line will soon be reviewed in connection with Department of Defense development of uniform standards for the evaluation of contractors' quality control organization procedures and techniques.

The second area covered existing contracts for which production was underway but not satisfactory in all respects. These contracts required and have been given the necessary detailed attention by the inspector on the job. In some cases, the unsatisfactory condition was due to delayed delivery of materials rather than quality of product. However, because unsatisfactory performance generally produced pressure on the contractor, these contracts were allowed to run their regular course to completion. It was considered inadvisable to introduce additional so called "complications".

The third area was that of new contracts that did not incorporate revised specifications or stipulations requiring the contractor to use quality control procedures in accordance with the new policy. These

contracts were considered in the light of the Inspector's past experience with the contractor. If the experience was satisfactory, then the attempt was made to sell the new quality control procedures without increase in cost. If the experience had been unsatisfactory, a determined effort was made to have the contractor install suitable measures for improving his quality, but full Government use of the contractor generated quality data was deferred until time and experience could test the reliability of this data.

The fourth area was that of new contracts that did contain provisions for quality control procedures with and without revised specifications. These contracts, while not as numerous as we would have liked, received detailed attention. We policed them very carefully with the contractors and arrived at agreed quality control procedures, records, tests and inspection standards. They are now proceeding satisfactorily.

These were the areas and they encompassed many operational "stickers" which could not be avoided nor readily eliminated. Now we 11 discuss areas in which Navy can show factual progress.

Contracting Officers were encouraged to include the new quality control provisions in their contracts. While the year 1958 did not show as much progress along these lines as desired, firm steps are being taken in 1959 to insure that the new contracts incorporate the required quality control provisions. Contracting Officers are advised, through preaward surveys and through communication with the Inspector, of marginal producers who are not considered capable of producing satisfactory materials. This is important because the screening out of unsatisfactory producers before they receive contracts is one of the best ways of assuring a quality product. Because of the rules imposed on Contracting Officers concerning awarding to the lowest bidder, helping distressed labor areas and patronizing small business, it is not always possible to avoid contracting with a marginal producer.

Progress was made in revising specifications to reflect the contractor responsibility policy on tests and inspections. Progress along these lines has been slow and not entirely satisfactory, some specifications revised during 1958 for other reasons, did not incorporate the changes necessary to support the policy. More determined efforts are now being made to bring specifications into conformance with policy.

Navy does not expect and will not require that all contractors, large and small, have a formal quality control department complete with statisticians, process controls and the like. A large number of contractors are small producers and employ few people. With others, the portion of the output for the Military is so small when compared to that for commercial customers that we would be presumptious to demand that they change their quality control methods. Not all of our products are as complicated as the POLARIS missile or the propulsion plant of NAUTILUS. Many can be tested adequately and measured after completion. Even with these, we want the producer to test and measure them before he presents them to our inspectors for acceptance.

At the field inspection level, the policy on quality assurance through surveillance inspection has resulted, during 1958, in revised personnel planning. Lower inspection positions, in which the incumbent was a line inspector doing routine dimensional and visual inspecting, were replaced by higher level positions where the incumbent was to evaluate methods, procedures, and contractor-generated data. The attrition, normal and induced, in positions GS-3 - GS-6 has been approximately 64%. Increase at the GS-8 level has been approximately 25%. A part of this attrition was due to reduction in workload as a result of fewer contracts. Most of it came about as a result of less lower level inspection on existing contracts. In connection with the change of surveillance inspection, a new job sheet for the field inspector was devised and is now being evaluated in service.

As a result of increased complexity and demand for reliability in modern weapon and propulsion systems, more intensive training has been provided for our quality control personnel. A special school was established for those involved with nuclear propulsion. Prospective students were nominated by the Officers in Charge of our various inspection offices. Then each nominee was interviewed for aptitude by a specialist in nuclear propulsion. The selected students were given six weeks of intensive classroom and laboratory training, with emphasis on nondestructive testing. This was followed by six months of on-the-job training, with an Atomic Energy Commission specialist looking over their shoulders, before they were turned loose to inspect nuclear cores on their own.

Progress was made in 1958 in sending to the field proposed procedures for evaluating the contractors' quality control methods. There were, in addition to the formal procedures of the technical bureaus, informal procedures in the various inspection offices that had been in use for years in handling inspections on a limited basis. No over-all Navy coordinated system of evaluation has as yet been formalized as it had been determined that a joint Army-Navy-Air Force task group under Department of Defense sponsorship will develop a system applicable to all the Armed Services. When this system is promulgated, another school will be established to train inspectors in methods of evaluating contractors' quality control systems and in making optimal use of contractor-generated data. Maximum national security for our defense dollar demands the finest teamwork between Industry and Government. We intend to improve this teamwork.

THE MIXED WEIBULL DISTRIBUTION IN LIFE-TESTING OF ELECTRON TUBES*

John H. K. Kao
Cornell University

Introduction and Summary:

The use of Weibull distribution (1) in characterizing the lifetime of electron tubes was first studied by the author some eight years ago. Some of the results were given in (2,3,4,5). The present paper includes a refinement in the mathematical model as follows. Instead of a single distribution, a mixture of two Weibull distributions each representing a type of tube failure is postulated. The concept of two failure types-sudden and delayed-concurs with the field experience on tube failure. This new model proposed here, allows even more flexibility in data fitting with very little additional complexity over the single-population model used earlier. A simple graphical method of estimating the parameters of the mixed Weibull distribution on a special graph paper is illustrated by using actual life-testing data gathered by a small-scaled life-test of some 800 6AQ5A's conducted at Cornell University under a Signal Corps contract. The construction of a statistical confidence band on the theoretical failure distribution utilizing the incomplete beta function table is also described.

Classification of Failure Types.

In broad terms electron tube failures can be classified into two types:

Catastrophic or Sudden filures and
Wear-out or Delayed filures.

A tube which is operating normally and then suddenly becomes completely inoperative is defined here as a catastrophic failure. A tube whose characteristics gradually fall outside of some specifications is defined as a wear-out failure.

1. Catastrophic or Sudden Failures. In general, the most important causes for catastrophic failure are of a mechanical nature. The catastrophic failures include glass failure, short circuits and open circuits. The glass failures are mainly due to strain cracks in the glass, and seal leaks; the shorts are due to loose conducting particles, such as lint, and to actual contact of parts that were spaced too close, such as pushed grid-wire turns; and the opens are due to open welds or burned out elements. Often, there are also indirect causes. For example, the current in a tube which is a slow leaker may rise excessively and cause a tube element (e.g. screen grid) to melt and short to another element.

For most applications, since it is usually the poor-quality tubes (the weaker sisters) that fail prematurely, the hazard rate (to be

*Work done under U.S. Signal Corps Contract DA 36-039-SC-64646 with Cornell University.

defined later) of the population decreases with operation on life. Furthermore, it is reasonable to assume that failures of these poor-quality tubes would actually start as soon as they are exposed to risks. Under these considerations a Weibull distribution with location parameter equal to zero and shape parameter less than one would be an appropriate mathematical model.

2. Wear-out or Delayed Failures. The most important causes for wear-out failure are of an electrical or chemical nature. The wear-out failure is associated with the "wearing-out" of one or more of tube's elements. In the case of a conventional receiving tube, to cite a specific example, the cathode element "wears-out" as a result of the evaporation of barium from the coating, the loss of coating and barium due to ion bombardment, the formation of interface resistance, and other related factors. The "wearing-out" of the cathode effects such important tube characteristics as transconductance, peak emission, plate current, etc. which often have specific life-test end-points. In actual field use of tubes, the slumping of tube characteristics causes "wear-out" failure as according to applications which include circuit requirement such as various duty cycle, environmental stress such as elevated temperature, maintenance policy such as marginal checking, legal regulations such as those imposed on airborn applications, etc.

For most applications, very few wear-out failures occur until some particular time period has elapsed after which the slumping of characteristics causes the hazard rate to increase with time. The length of this delay period depends upon application and in the case of static life-testing upon the specific definition of life-test end-point. In view of these considerations a Weibull distribution with some positive location parameter and shape parameter larger than unity seems to be a reasonable model.

Combining the above two failure causes a mixture of two Weibull populations as described in the following sections is postulated as the underlying distribution of the life time of electron tubes.

Definitions.

We give the following definitions which are relevant to this paper.

1. Simple Weibull Distribution. A simple Weibull cumulative distribution function (c.d.f.) is defined as,

$$(1) \qquad F(x) = 1 - e^{-(x-\gamma)^{\beta}/\alpha} \quad \text{for } x \geq \gamma \,;\, \alpha, \beta \text{ positive.}$$

α = scale parameter
β = shape parameter
γ = location parameter*

The first derivative of $F(x)$ with respect to x is called a simple Weibull probability density function (p.d.f.), thus,

$$(2) \qquad dF(x)/dx = f(x) = \frac{\beta(x-\gamma)^{\beta-1}}{\alpha} \; e^{-(x-\gamma)^{\beta}/\alpha}$$

*The notations originally used by Weibull(1) are: x_o, m and x_u which

Figures (1) and (2) show some of the simple Weibull distributions.

Fig. Simple Weibull c.d.f. ($\alpha = 1$)

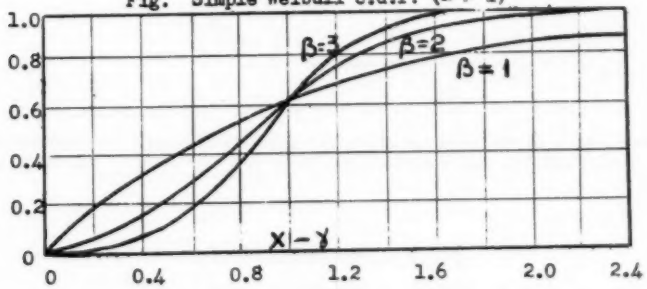
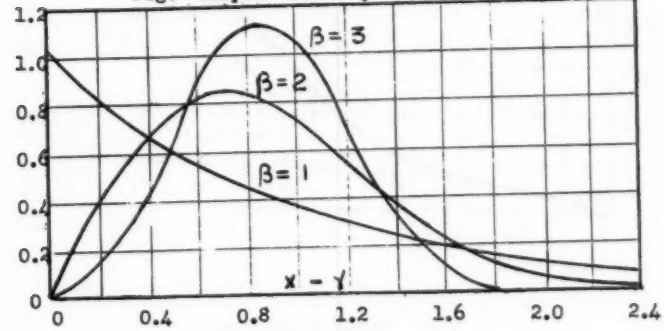


Fig. Simple Weibull p.d.f. ($\alpha = 1$)



2. Mixed Weibull Distribution. A k-fold mixed Weibull c.d.f. is defined as,

(3) $$F(x) = \sum_{i=1}^{k} p_i F_i(x) \quad \text{where } 0 \leq p_i \leq 1,$$

$$\sum_{i=1}^{k} p_i = 1 \text{ and } F_i(x) = 1 - e^{-(x-\gamma_i)^{\beta_i/\alpha_i}} \text{ is called the ith}$$

sub-population in c.d.f. form. The quantities p_i's are the proportions of sub-population mix or simply mix parameters. If we put $R_i(x) = 1 - F_i(x)$ then

(4) $$F(x) = 1 - \sum_{i=1}^{k} p_i R_i(x) \quad .$$

---

correspond to our α, β, and γ respectively.

A k-fold mixed Weibull p.d.f. is therefore,

$$(5) \qquad f(x) = \sum_{i=1}^{k} p_i f_i(x) \text{ with sub-population in p.d.f. form}$$

$$f_i(x) = \frac{\beta_i (x - \gamma_i)^{\beta_i - 1}}{\alpha_i} \, e^{-(x-\gamma_i)^{\beta_i}/\alpha_i}$$

Figures 3 and 4 show a 2-fold mixed Weibull c.d.f. and its corresponding p.d.f. each with two sub-populations with mix parameters equal to p and $1 - p = q$.

Fig.3 Mixed Weibull c.d.f.

$\alpha_1 = \alpha_2 = 1 \quad \beta_1 = 1/2 \quad \beta_2 = 3 \quad p = .20 \quad \gamma_1 = 0 \quad \gamma_2 = .40$

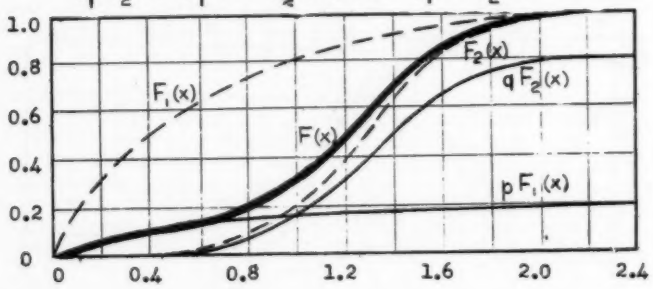
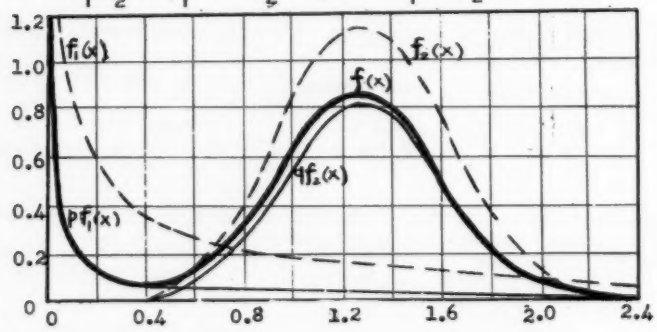


Fig.4 Mixed Weibull p.d.f.

$\alpha_1 = \alpha_2 = 1 \quad \beta_1 = 1/2 \quad \beta_2 = 3 \quad p = .20 \quad \gamma_1 = 0 \quad \gamma_2 = .40$



3. Composite Weibull Distribution. A r-component composite Weibull c.d.f. is defined as,

$$(6) \qquad F(x) = F_j(x) \text{ for } \delta_j \leq x \leq \delta_{j+1}, \quad j = 0, 1, 2, \ldots r;$$

where $F_j(x) = 1 - e^{-(x-\gamma_j)^{\beta_j} / \alpha_j}$ is called the jth component in c.d.f. form. The quantities δ_j's are the points of component partition or simply partition parameters. Since the end points are $\delta_o = \gamma_1$ and $\delta_{r+1} = \infty$, hence for a r-component composite Weibull distribution there are only (r-1) partition parameters.

A r-component Weibull p.d.f. is

$$(7) \qquad f(x) = f_j(x), \text{ for } \delta_j \leq x \leq \delta_{j+1}, \quad j = 0, 1, 2, \ldots, r;$$

where the jth component is $f_j(x) = \dfrac{\beta_j (x-\gamma_j)^{\beta_j - 1}}{\alpha_j} e^{-(x-\gamma_j)^{\beta_j}/\alpha_j}$

Figures 5 and 6 show a 2- component composite Weibull c.d.f. and its corresponding p.d.f. partitioned at δ.

Fig. 5 Composite Weibull p.d.f.

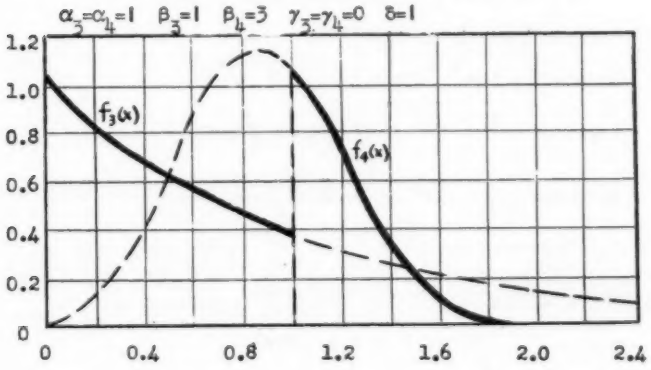
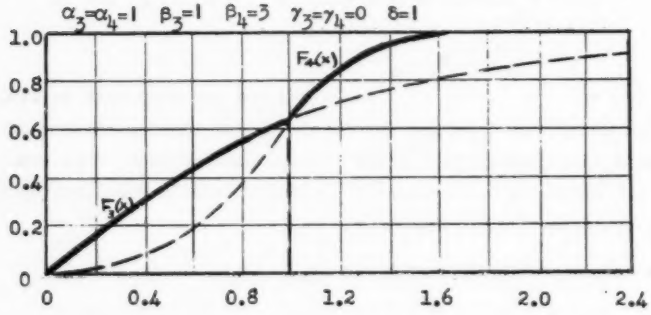


Fig. 6 Composite Weibull c.d.f.



4. Moments of Lifetime Distribution. Given $f(x)$ and $F(x)$ as the p.d.f. and c.d.f. of lifetime with $\gamma \leq x < \infty$, by a theorem shown in Appendix B the kth moment is

$$(8) \qquad \mu'_k = \int_{-\infty}^{\infty} x^k f(x)\,dx = \gamma^k + k \int_{\gamma}^{\infty} x^{k-1}[1-F(x)]\,dx$$

Equation (8) simplifies the computation of moments of Weibull distributions considerably. If we put

$$b = \frac{1}{\beta} \quad \text{and} \quad \Gamma(y) = \int_0^{\infty} z^{y-1} e^{-z}\,dz \ ,$$

then for a simple Weibull distribution,

$$(9) \qquad \text{Mean } \mu = \gamma + \alpha^b\, \Gamma(b+1)$$

$$(10) \qquad \text{Variance } \sigma^2 = \alpha^{2b} \{ \Gamma(2b+1) - \Gamma^2(b+1) \}$$

$$(11) \qquad \text{Skewness } \alpha_3 = \frac{\Gamma(3b+1) - 3\Gamma(2b+1)\Gamma(b+1) + 2\Gamma^3(b+1)}{\{\Gamma(2b+1) - \Gamma^2(b+1)\}^{3/2}}$$

5. Survival Function and Reliable Life. Survival function $R(x)$ which is also known as reliability function is defined as,

$$(12) \qquad R(x) = \int_x^{\infty} f(y)\,dy = 1 - F(x)$$

The reliable life, ρ_R, which is sometimes called the minimum life, is defined for any specified R such that,

$$(14) \qquad R = \int_{\rho}^{\infty} f(y)\,dy$$

The reliable life ρ_R is the same as the quantile of order p if p=1-R.

A special case is when $R = \frac{1}{2}$, then ρ becomes the median ξ. In the case of a simple Weibull distribution, we find

(15) Weibull $R(x) = e^{-\frac{(x-\gamma)^\beta}{\alpha}}$ and

(16) Weibull $\rho_R = \gamma + \alpha^b(-\ln R)^b$ where lnx denotes the natural logarithm of x.

(17) Weibull $\xi = \gamma + \alpha^b(\ln 2)^b$.

6. Failure Rate, Hazard Rate and Retired Life. Failure rate for period of length h, denoted by G(x,h) is defined as,

(18) $G(x,h) = \frac{1}{h}\int_x^{x+h} \{ f(y)/R(x)\}\, dy = \frac{1}{hR(x)}\{ F(x+h) - F(x)\}$

The last member of the equation is obtained by the definition of F(x). Hazard rate, Z(x), also known as instantaneous failure rate, is defined as,

(19) $Z(x) = \lim_{h\to 0} G(x,h) = f(x)/R(x)$.

The last member of the equation is obtained by the definition of f(x). Retired life, ζ_z, which is sometimes called replacement life, is defined for any specified Z such that,

$$Z = f(\zeta)/R(\zeta)$$

In the case of a simple Weibull distribution, we obtain,

(20) Weibull $G(x,h) = \frac{1}{h}\{1 - e^{\frac{(x-\gamma)^\beta - (x-\gamma+h)^\beta}{\alpha}}\}$

(21) Weibull $Z(x) = \beta(x-\gamma)^{\beta-1}/\alpha$

(22) Weibull $\zeta_z = \gamma + (\frac{\alpha z}{\beta})^{\frac{1}{\beta-1}}$

7. Life Expectancy. The life expectancy, denoted by L(x), which is the additional expected life of an item aged x, is defined as,

(23a) For $x \le \gamma$, $L(x) = \{\gamma - x + \int^\infty R(t)dt\} / R(\gamma) = \gamma - x + \mu$.

(23b) For $x > \gamma$, $L(x) = \int_x^\infty R(t)dt / R(x) = \int_0^\infty R(x+y)dy / R(x)$.

For a simple Weibull distribution, we have for the second case,

(24) Weibull $L(x) = \frac{b\,\alpha^b}{R(x)}\ \Gamma\left(b \mid \frac{(x-\gamma)^\beta}{\alpha}\right)$*

8. Probable Life. The probable life, denoted by B(x), which is the total expected life of an item aged x, is defined as,

(25) $B(x) = L(x) + x.$

It can be seen that for a new item, i.e. $x \leq \gamma$ both L(x) and B(x) become the same as mean lifeμ. Hence, Equations (23), (24), and (25) are usually used as measures of life quality when the items are not new.

The Proposed Model-Mixed Weibull Distribution.

The foregoing discussions on catastrophic and wear-out failure types of electron tubes suggest the following two sub-populations:

For catastrophic or sudden failures,

(26) $F_1(x) = 1 - e^{-x^{\beta_1}/\alpha_1}$, for $x > 0$ and $\alpha_1 > 0$, $0 < \beta_1 < 1$.

For wear-out or delayed failures,

(27) $F_2(x) = 1 - e^{-(x-\gamma_2)^{\beta_2}/\alpha_2}$, for $x > \gamma_2$ and $\alpha_2 > 0, \beta_2 > 1$.

A mixture of these two sub-populations is our proposed model, thus,

(28) $F(x) = p\,F_1(x) + q\,F_2(x) = 1 - p\,e^{-x^{\beta_1}/\alpha_1} - q\,e^{-(x-\gamma_2)^{\beta_2}/\alpha_2}$

with $p+q = 1$. Equation (28) states that at a given application the life time of an electron tube is a mixed chance variable depending not only on the tube's mechanical strength but also on the tube's wear-out quality. The catastrophic scale and shape parameters α_1 and β_1; the wear-out scale and shape parameters α_2 and β_2; the delay period γ_2 and mix proportion p are all to be estimated from the life-testing or field failure data. The following graphical method of estimating these parameters is based upon the fact that a simple Weibull cumulative distribution becomes a straight line in ln versus ln-ln coordinates. (Refer to the example given in Table I and Figure 7).

1. Plot the sample c.d.f. of the life-testing or field failure data on the Weibull probability paper and fit by inspection a curve (called Weibull plot) among these points. **

---

* $\Gamma(K \mid a) = \int_0^a y^{K-1} e^{-y} dy$ which is called the incomplete gama function and is tabulated in (8).

** For small sample size, it is better to use $\frac{j}{n+1}$ instead of j/n for the jth sample c.d.f. See Appendix A for details.

** Weibull probability paper contains ln versus ln-ln scales and other scales calibrated for life-testing use. It is available from Cornell University, Ithaca, New York

2. Starting from each end of the Weibull plot draw a tangent line and denote them by $\widehat{pF}_1$ and $\widehat{qF}_2$ which are estimates of $pF_1(x)$ and $qF_2(x)$ respectively.

3. The intersection of $\widehat{qF}_2$ with the bottom borderline gives $\hat{\gamma}_2$ which estimates γ_2.

4. At the intersection of $\widehat{qF}_2$ with the upper borderline drop a vertical line whose intersection with pF_1 as read from the Percent Failure scale gives $\hat{p}$ which estimates p (and hence $q = 1-p$).

5. Knowing the estimate of p, the mixed Weibull population may now be separated into the two sub-populations. The estimated number of items belong to $F_1(x)$ and $F_2(x)$ are $n\hat{p}$ and $n(1-\hat{p})$ or $n\hat{q}$ respectively from which the sample c.d.f.'s of $F_1(x)$ and $F_2(x)$ may be calculated.

6. The y-intercept and slope of the Weibull plots of the separated sub-populations Fi(x), i = 1&2, give the usual estimates of $\ln\alpha_i$ and β_i respectively.

An Approximation to the Proposed Model-Composite Weibull Distribution.

The above graphical procedure of estimation seems quite involved. A somewhat simpler method is available if the mixed model can be approximated by a composite model. This approximation is good for small p and large γ_2 in the mixed model. In this case, the Weibull plot would consist of two distinct linear portions which can be treated as the two components of a composite population as follows:

$$(29)\quad F(x) = F_3(x) = 1 - e^{-x^{\beta_3}/\alpha_3},\ \text{for } 0 \le x \le \delta,\ \alpha_3 > 0,\ 0 < \beta_3 < 1;$$

$$F(x) = F_4(x) = 1 - e^{-x^{\beta_4}/\alpha_4},\ \text{for } \delta \le x \le \infty,\ \alpha_4 > 0,\ \beta_4 > 1.$$

Equation (29) is a 2-component Weibull c.d.f. with both location parameters equal to zero, catastrophic scale and shape parameters equal to α_3, β_3 and wear-out scale and shape parameters equal to α_4, β_4.* The partition parameter δ at which the two components intersect is not a real parameter, because it can be expressed in terms of other parameters. At time δ the proportion of catastrophic failure is equal to that of wear out failure, thus,

$$(30)\quad 1 - e^{-\delta^{\beta_3}/\alpha_3} = 1 - e^{-\delta^{\beta_4}/\alpha_4},\ \text{hence}$$

$$= \left(\frac{\alpha_4}{\alpha_3}\right)^{\frac{1}{\beta_4 - \beta_3}} = \exp\left(\frac{\ln\alpha_4 - \ln\alpha_3}{\beta_4 - \beta_3}\right)$$

---

* The subscripts 3 and 4 used here are to distringuish from the subscripts 1 and 2 already used for the mixed model.

Estimation of parameters in the composite case becomes much simpler. (Refer to the example given in Table I and Figure 7).

1. Plot the sample c.d.f. (See Appendix A for small samples) of the life-testing or field-failure data on the Weibull probability paper and fit by inspection two intersecting straight lines among the points.

2. The intersection of these two components as read from the Failure-Age scale gives the estimate for δ .

3. The y-intercept and slope of these two components gives the estimates of $\ln \alpha_i$ and β_i, i = 3 and 4.

Confidence Band for F(x).

Just as the Weibull plot provides a point-by-point estimates of F(x), the following confidence band will give an interval-by-interval estimates of F(x). This confidence band can be found as follows.

In Appendix A, it is shown that for a set of ordered observations denoted by $x_1 \leq x_2 \leq \ldots \leq x_n$ which were originally independent and identically distributed, the probability transformation $Y_i = F(X_i)$ for all i = 1, 2, ...,n would give the p.d.f. of Yi as the following beta distribution:

$$(31) \qquad h(y_i) = \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\, y_i^{a-1} (1-y_i)^{b-1} \quad \text{for } 0 \leq y_i \leq 1,$$

where a=i and b=n+1-i.
The c.d.f. of Yi denoted by $H(y_i)$ is the so-called incomplete beta function and is given in graphical form in Table 7 of (6) as $I_{y_i}(a, b)$ which is of course identical with $H(y_i)$. Knowing any three of these four quantities: I, y_i, a and b, the remaining one can be easily found by utilizing the following identity which was implied by Forms 1 and 2 of $H(y_i)$ derived in Appendix A.

$$(32) \qquad I_{1-y_i}(b,a) = 1 - I_{y_i}(a,b)$$

In Appendix A, $H(y_i) = I_{y_i}(i, n+1-i)$ is equated to one half for finding the medians of Y_i. By the same token we can set $H(y_i)$ equal to any number between zero and one and obtain the so-called quantile of Y_i. Hence for the confidence coefficient of $100(1-\alpha)\%$, the lower confidence limit $L(y_i)$ and the upper confidence limit $U(y_i)$ are given by

$$(33) \qquad L(y_i) = H_{y_i}^{-1}(\alpha/2) \quad \text{and}$$

$$U(y_i) = H_{y_i}^{-1}(1-\alpha/2) \quad \text{for } i = 1, 2, \ldots, n.$$

Plotting (33) with each xi we obtained the $100(1-\alpha)\%$ confidence band for $F(x_i)$.

To illustrate, take n=10, i=2, the lower confidence limit of the

TABLE I

A Summary of Cornell Experiment, 400 Type 6AQ5A, Wear-Out Failure Criterion: 73.2% g_m Drop.

	Estimated failure lifetime in hrs.	Sample c.d.f in %.
a. Mixed populations: (n = 400)		
1	19.5	.25
2	84.4	.50
3	362	.75
4	488	1.00
5	1050	1.25
6	1390	1.50
7	1650	2.75
8	2060	4.75
9	2390	8.50
10	2650	19.00
11	3130	32.75
12	3810	56.00
13	4530	75.50
14	5060	91.75
b. Sub-population I, $F_1(x)$: ($n\hat{p}$ = 400(.023) = 9.2, say 10)		
1	19.5	10
2	84.4	20
3	362	30
4	488	40
5	1050	50
6	2390	60

c. Sub-population II, $F_2(x)$: ($n\hat{q}$ = 400-$n\hat{p}$ = 390, $\hat{\gamma}_2$ = 800)

	x	x - γ_2	
1	(1390)	590	.256
2	(1650)	850	1.538
3	(2060)	1260	3.590
4	(2390)	1590	7.179
5	(2650)	1850	17.949
6	(3130)	2330	32.051
7	(3810)	3010	55.897
8	(4530)	3730	75.897
9	(5060)	4260	92.564

NOTES:

1. The six points circled are catastrophic failures.
2. For sample size as large as 400 used here, the sample c.d.f., the sample mean of $F(X_1)$ and the sample median of $F(X_1)$ yield almost the same value, hence only sample c.d.f.'s are shown. For $F_1(x)$ see Table II.

TABLE II

Confidence Band (80%) for Catastrophic Sub-population, $F_1(x)$

	Time readings last inspection in hrs.	Time readings this inspection in hrs.	Est'd failure lifetime in hrs.	Sample c.d.f. in %	Sample Median in %	Upper Confidence Limit in %	Lower Confidence Limit in %
1	8.5	30.5	19.5	10	6.7	1.0	20
2	72.2	96.7	84.4	20	16.2	5.5	33
3	309.6	414.3	362	30	25.9	11	44
4	414.3	562.5	488	40	35.5	18	55
5	1001	1100	1050	50	45.2	27	64
6	2375	2411	2390	60	54.8	36	73

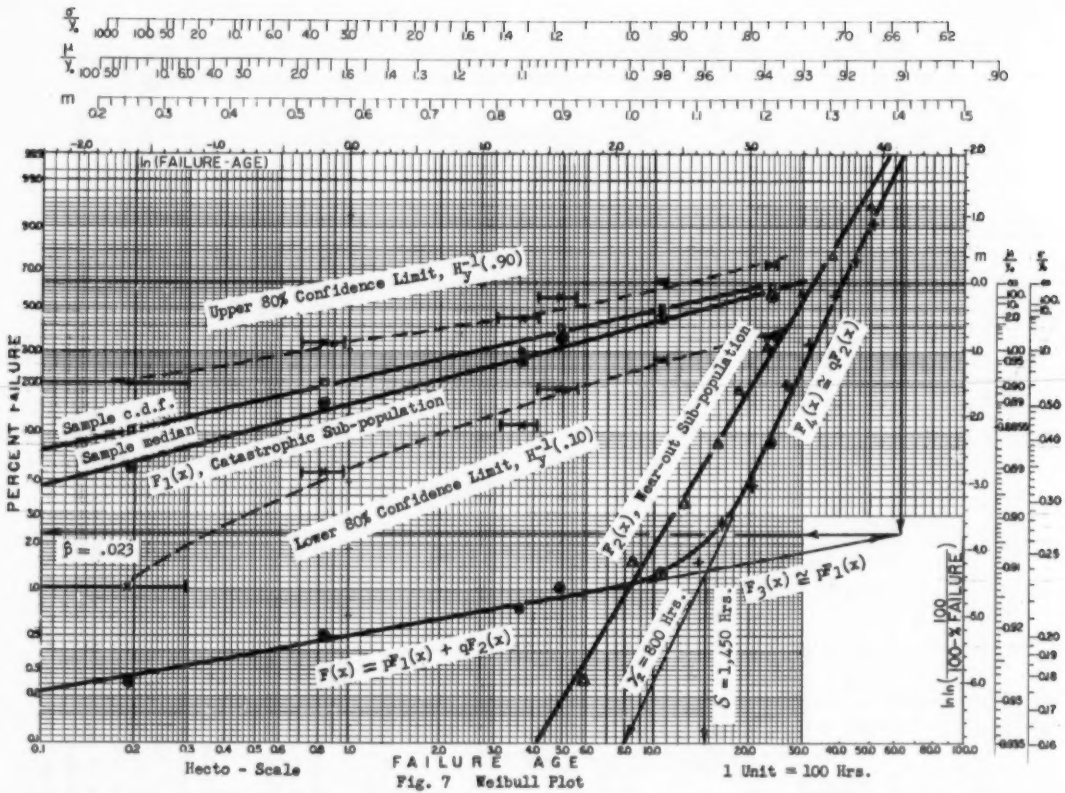


Fig. 7 Weibull Plot

2-sided 80% confidence interval for $F(x_2)$ is $L(y_2) = H_2^{-1}(.10)$. This is equivalent to $I_{y_2}(2,8) = .10$. But this expression can not be found in the incomplete beta function table, hence we make use of the identity given by Equation (32) by stating $I_{1-y_2}(8,2) = .90$ and solve for $(1-y_2)$. This is done by referring to Table 17 of (6) which gives $(1-y_2) = .945$ or $y_2 = .055$. (See the second entry on the upper confidence limit in Table II).

An Example.

In Table I, the section (a) shows the abridged version of the life-testing data obtained from a Cornell experiment (7) on 800 Type 6AQ5A conducted under Signal Corps Contract DA36-039-sc-64646.

Figure 7 shows the Weibull plot. In this case, the estimate of $p=2.3\%$(small) and $\gamma_2 = 800$ hrs. (large), hence the catastrophic and wear-out components of the failure distribution are quite pronounce. The sections (b) and (c) of Table I give the Weibull plots of $F_1(x)$ and $F_2(x)$ which are shown in Figure 7. The following are estimates of the remaining parameters:

α_1	α_2	α_3	α_4	β_1	β_2	β_3	β_4	δ
4.48	1.1×10^5	200	5.4×10^6	.43	3.3	.36	4.2	14.50

A two-sided 80% confidence band of $F_1(x)$ is given in Table II and shown in Figure 7 together with the median of $F_1(x)$.

Figure 8 shows a plot of the hazard rate for the Cornell data using the estimates of parameters based upon the composite Weibull distribution. That is,

$$(34) \qquad Z(x) = \frac{.36\, x^{-.64}}{200} \quad \text{for } x < 1450 \text{ hrs.}$$

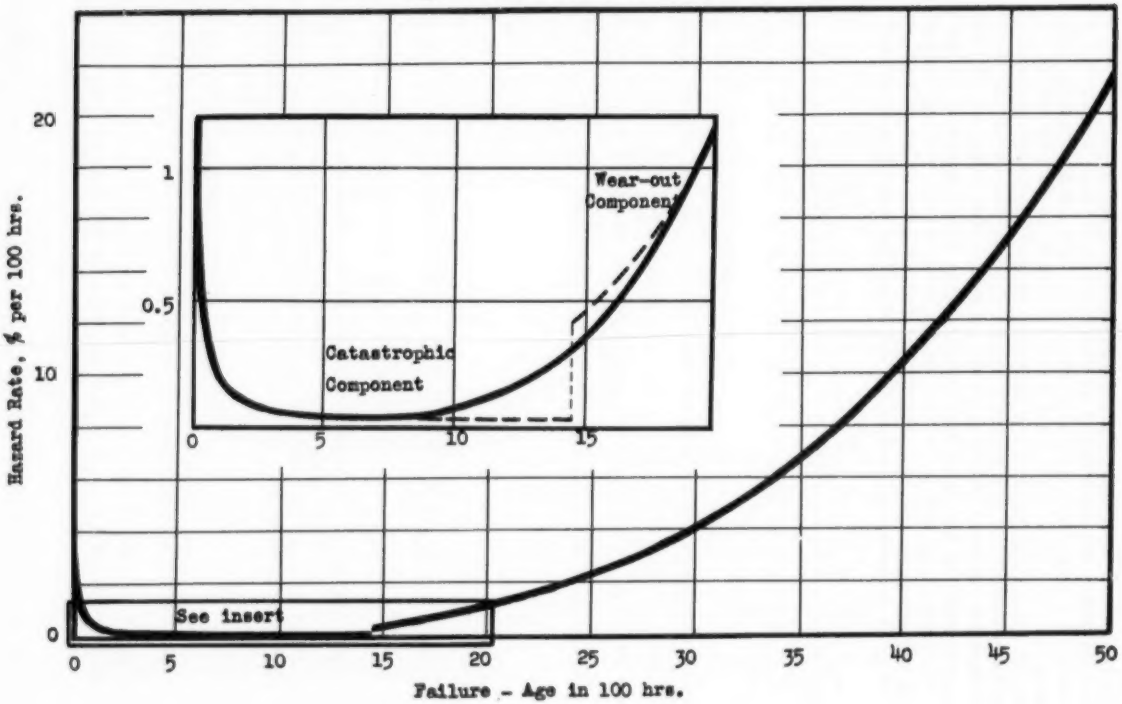
$$= \frac{4.2\, x^{3.2}}{5{,}400{,}000} \quad \text{for } x > 1450 \text{ hrs.}$$

Some Remarks.

If the Cornell data are representative, then it can be seen that before the partition time (1450 hrs for Cornell data), the hazard rate decreases with time and after that the hazard rate sharply increases. This empieical fact was observed by many experimenters on tube failures.

The discontinuity near the partition point should be smoothed out, since the composite Weibull distribution is an approximation to the proposed mixed model. The insert in Figure 8 showed the enlarged portion of the lower left corner of Figure 8. Note that for a brief time-period (from approximately 500 hrs. to 1000 hrs.) the hazard rate stabilizes to a constant value, hence for this time-period the failures can be considered as occuring randomly and the exponential distribution applies. However, this period would be shortened if the circuit requirement imposed on the tubes becomes more stringent.

Fig. 8. Hazard Rate for Cornell Data



It is interesting to note that if the Cornell test were stopped at 500 hrs., as practiced by tube manufacturers on many military and commercial tubes, then the tubes removed from the life-testing racks would be better than the original new tubes. This is curious especially when life-tests are considered as destructive tests. Hence, the widely practiced 48 hour "burn-in" as part of the tube manufacturing process in weeding out the so-called "weak sisters" seems to be reasonable. For tubes on missile application a somewhat longer "burn-in" seems to be even better provided the circuit requirement is not too critical as to move forward the wear-out component too much. For other areas, such as the industrial or computer application where tube longevity is important, the "burn-in" only has the effect of shortening the "de-bug" period and in no way prolongs the replacement or retired life of electron tubes. In fact, it consumes a portion of the otherwise useful life of the tube.

APPENDIX A

THREE ESTIMATES OF THE F(X) POPULATION CUMULATIVE DISTRIBUTION FUNCTION (c.d.f.)

1. Sample or Empiric c.d.f.

1. Ungrouped data: From a life-test of sample size n, let $x_1 \leq x_2 \leq \cdots \leq x_r$ be the r exact failure-ages observed. The test is stopped when the rth failure appears where $r \leq n$. Let $\hat{F}(x_i)$ be the estimate of $F(x_i)$, then the sample c.d.f., i.e., the proportion failure on or before x_i is given by:

$$\hat{F}(x_i) = i/n$$

If $r = n$, i.e., the complete data, then $\hat{F}(x_n) = 1$. Since in the Weibull plot the quantity $\ln\ln \frac{1}{1-\hat{F}(x_i)}$ is plotted as the principle ordinate, we would have to sacrifice the last point because $\ln\ln \frac{1}{1-\hat{F}(x_n)} = \infty$ which obviously can not be plotted.

2. Grouped data: Let the failure-data be grouped into: $z_1, f_1; z_2, f_2; \cdots; z_k, f_k$ where z_j is the class-mark (a middle point) of the jth grouping cell and f_j is the corresponding frequency of failures in the jth cell with $\sum_{j=1}^{k} f_j \leq n$. Then the estimated c.d.f. for ith cell is:

$$F(z_i) = \sum_{j=1}^{i} f_j/n.$$

If we have the complete data, then $\sum_{j=1}^{k} = n$ & $\hat{F}(z_k) = 1$. Again, the last point is omitted in the Weibull plot.

Accordingly, for complete life-testing data if the sample size is small (if the number of cells is small in the case where grouped data are used) the information represented by the largest cell is sacrificed in the Weibull plot. This disadvantage can be rectified if the next method of getting $\hat{F}(x)$ is used.

II. Sample Mean of $F(X_i)$.

Let $x_1 \leq x_2 \leq \cdots \leq x_n$ be a set of n independent observations ordered according to their size from a p.d.f. f(x) whose c.d.f. is F(x). Then by the multinomial theorem for $p_1 + p_2 + p_3 = 1$ the probability element $g(x_i)dx_i$ is:

$$\frac{n!}{(i-1)!\,1\,(n-i)!} p_1^{i-1} p_2^{1} p_3^{n-i} = C_n^{i} [F(x_i)]^{i-1} [1-F(x_i)]^{n-i} f(x_i)dx_i$$

where $C_n = n!/(i-1)!\,(n-i)!$. Allowing a change of variable $y_i = F(x_i)$ or $dy_i = f(x_i)dx_i$ the p.d.f. of y_i is,

$$h(y_i) = C_n\, y_i^{i-1} (1-y_i)^{n-i} \quad \text{for } 0 \leq y_i \leq 1,$$

since $F(x_i) = F\{F^{-1}(y_i)\} = y_i$ and $h(y_i) = g\{F^{-1}(y_i)\} \cdot |J|$ where J is the Jacobian of the transformation and,

$$J = dx_i/dy_i = 1/\{dF(x_i)/dx_i\} = 1/f(x_i).$$

Since n and i are integers, we may write C_n as $\Gamma(n+1)/\{\Gamma(i)\ \Gamma(n+1-i)\}$.

The density h(y) can now be recognized as a beta distribution with parameters: a= i and b = n+1-i, where the beta distribution (p.d.f.) is defined as:

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\ \Gamma(b)}\ x^{a-1}(1-x)^{b-1}, \text{ for } 0 \leq x \leq 1,\ a,b > 0$$

From the identity that $a!\ b! = (a+b+1)! \int_0^1 x^a(1-x)^b dx$, the kth moment of y_i about the origin is,

$$E(y_i^k) = C_n \int_0^1 y^{k+i-1}(1-y)^{n-i}dy = \frac{n!}{(i-1)!\ (n-i)!} \cdot \frac{(k+i-1)!(n-i)!}{(k+n)!}$$

$$= \frac{n!\ (k+i-1)!}{(i-1)!(k+n)!}.$$

Hence

$$E(y_i) = \frac{n!\ i!}{(i-1)!\ (n+1)!} = \frac{i}{n+1}.$$

Recall that $y_i = F(x_i)$, the expression above states that $i/(n+1)$ is an unbiased estimate of $F(x_i)$ hence for,

1. Ungrouped data: The sample mean of $F(X_i)$ is given as:

$$\hat{F}(x_i) = i/(n+1)$$

If i=n, i.e., the complete data, then $\hat{F}(x_n) = n/(n+1)$ which is still within graph for n less than 1000.

2. Grouped data: In this case, the sample mean of $F(X_i)$ is given by:

$$\hat{F}(z_i) = \sum_{j=1}^{i} f_j/(n+1), \text{ where } z_i, \text{ for each i, are defined as}$$

before.

gain if we have the complete data, then $\sum_{j=1}^{k} f_j = n$ and $\hat{F}(z_k) = n/(n+1)$ $= 1 - \frac{1}{n+1}$ and $\ln \ln \frac{1}{1-\hat{F}(z_k)} = \ln \ln (n+1)$ which is finite for any finite sample size n.

III. Sample Median of $F(X_i)$:

From $h(y_i)$, the p.d.f. of y_i derived above the c.d.f. of y_i, denoted by $H(y_i)$, may be found. The median of y_i is simply the solution of $H(y_i) = \frac{1}{2}$. The c.d.f. of y_i is,

$$H(y) = C_n \int_o^y z^{i-1}(1-z)^{n-i}dz = C_n \int_o^y z^{i-1} \sum_{j=o}^{n-i} (-1)^j \binom{n-i}{j} z^j dz$$

$$= C_n \sum_{j=o}^{n-i} (-1)^j \binom{n-i}{j} \int_o^y z^{j+i-1}dz = C_n \sum_{j=o}^{n-i} (-1)^j \binom{n-i}{j} y^{j+i} (j+i)^{-1}$$

(Form 1)

If we let $v = 1-z$, the above integration may be evaluated as:

$$H(y) = C_n \int_{1-y}^{1} (1-v)^{i-1} v^{n-i} dv = C_n \int_{1-y}^{1} v^{n-i} \sum_{j=o}^{i-1} (-1)^j \binom{i-1}{j} v^j dv$$

$$= C_n \sum_{j=o}^{i-1} (-1)^j \binom{i-1}{j} \int_{1-y}^{1} v^{j+n-i} dv$$

$$= C_n \sum_{j=o}^{i-1} (-1)^j \binom{i-1}{j} \left[1 - (1-y)^{j+n+1-i}\right] (j+n+1-i)^{-1}$$

(Form 2)

Using Form 2 of H(y), the setting of i=1 gives,

$$H(y_1) = 1 - (1-y_1)^n ,$$

hence

$$H(y_1) = \tfrac{1}{2} \text{ gives } \hat{y}_1 = 1 - (\tfrac{1}{2})^{\frac{1}{n}} = \hat{F}(x_1)$$

Using Form 1 of H(y), the setting of i = n gives,

$$H(y_n) = y_n^n ,$$

hence $H(y_n) = \frac{1}{2}$ gives $\hat{y} = (\frac{1}{2})^{\frac{1}{n}} = \hat{F}(x_n)$

These two explicit solutions give the estimates of $F(X_i)$ by its median at x_1 and x_n. For other values x_2, x_3 --- x_{n-1} (as a matter of fact x_1 and x_n also), the median may be found by setting $H(y_i) = \frac{1}{2}$ in Table 17 of (6) and solve for y_i.

APPENDIX B

MOMENTS OF A MIXED WEIBULL DISTRIBUTION

In order to find the population mean, variance, etc. of a mixed Weibull distribution, we shall prove a theorem which simplifies moment computations for many distributions for which $1-F(x)$ is in the closed form.

Theorem: Let X be a random variable with c.d.f. $= F(x)$ p.d.f. $= f(x)$ and reliability function, $R(x) = 1-F(x)$ all defined over $\gamma \leq x < \infty$ where γ is any real number, then,

$$\mu'_k = \int_{-\infty}^{\infty} x^k f(x)\,dx = \gamma^k + k\int_{\gamma}^{\infty} x^{k-1} R(x)dx$$

if $\lim_{x\to\infty} x^{k+1} f(x) = 0$ for any $k < \infty$.

Proof: Rewrite the last member of the above expression for μ'_k as follows:

$$\int_{\gamma}^{\infty}(1 - F(x))\, k\, x^{k-1}dx$$

Denote $u = 1-F(x)$ and $dv = k\, x^{k-1}dx$ and integrate this expression by parts. We get,

$$x^k \{ 1 - F(x) \} \Big|_{\gamma}^{\infty} + \mu'_k$$

The upper limit of the first term vanishes if $\lim_{x\to\infty} x^{k+1} f(x) = 0$ for any finite k and $F(\gamma) = 0$ by definition of c.d.f. Hence, the theorem is proved.

Corollary 1. If the distribution is defined over some finite range $\gamma \leq x \leq \delta$, then $F(\gamma) = 0$ $F(\delta) = 1$, the condition $\lim x^{k+1} f(x) = 0$ is unnecessary.

Corollary 2. If the distribution is defined over the range $0 \leq x < \infty$, then

$$\mu'_k = k\int_{o}^{\infty} x^{k-1} R(x)dx.$$

In the case of a mixed Weibull distribution $\lim_{x\to\infty} x^{k+1} f(x) = 0$, hence we can use the above theorem for finding the moments. The kth moment about the origin is,

$$\mu'_k = pk\int_{0}^{\infty} x^k\, e^{-x^{\beta_1}/\alpha_1}\, dx + q\gamma_2^{\,k} + qk\int_{\gamma}^{\infty} x^k\, e^{-(x-\gamma_2)^{\beta_2}/\alpha_2}\, dx$$

We proceed to evaluate the last term by letting $y = (x-\gamma_2)^{\beta_2}/\alpha_2$. Dropping the subscript "2" of parameters for ease of typing, the last term of μ'_k is,

$$qk\int_{\gamma}^{\infty} x^k e^{-(x-\gamma)^{\beta}/\alpha} dx = qk\int_{0}^{\infty} (\alpha^{\frac{1}{\beta}} y + \gamma)^{k-1} e^{-y\alpha^{\frac{1}{\beta}}} \beta^{-1} y^{\frac{2}{\beta}-1} dy$$

$$= qk\sum_{i=0}^{k-1} \binom{k-1}{i} \gamma^i \alpha^{\frac{k-i}{\beta}} \beta^{-1}\Gamma(\tfrac{k-i}{\beta})$$

The first term in μ_k' is found similarly. Hence,

$$\mu_k' = p\alpha_1^{\frac{k}{\beta_1}} \Gamma(\tfrac{k}{\beta_1} + 1) + q\gamma_2^k + qk \sum_{i=0}^{k-1} \binom{k-1}{i} \gamma_2^i \alpha_2^{\frac{k-i}{\beta_2}} \beta_2^{-1} \Gamma(\tfrac{k-i}{\beta_2})$$

By setting k = 1 in the above expression, we get

$$\mu = p\alpha_1^{\frac{1}{\beta_1}} \Gamma(\tfrac{1}{\beta_1} + 1) + q[\gamma_2 + \alpha_2^{\frac{1}{\beta_2}} \Gamma(\tfrac{1}{\beta_2} + 1)]$$

which is in fact the weighted average of the means of the sub-populations in accordance to the mixed parameters p and q. The higher moments, such as variance and skewness, have more complicated relationships although tedious but can be evaluated similarly.

REFERENCES

1. Weibull, W., "A Statistical Distribution Function of Wide Applicability," Journal of Applied Mechanics, pp. 293-7, V 18, 1951.

2. Kao, J. H. K., "The Weibull Distribution in Life-Testing of Electron Tubes," (Abstract), Journal of the American Statistical Association p. 514, V 51, n.275.

3. Kao, J. H. K., "A New Life-Quality Measure for Electron Tubes," Institute of Radio Engineers Transactions on Reliability and Quality Control, pp. 1-11, PGRQC-7, April 1957.

4. Kao, J. H. K., "Computer Methods for Estimating Weibull Parameters in Reliability Studies," Institute of Radio Engineers Transactions on Reliability and Quality Control, pp. 15-22, PGRQC-13, July, 1958.

5. Kao, J. H. K., "The Design and Analysis of Life-Testing Experiments", American Society for Quality Control 1958 Middle Atlantic Conference Transactions, pp. 217-33, 1958.

6. Pearson, E. S. and Hartley, H. O., (Editors), Biometrika Tables for Statistians, V. I, Cambridge University Press, 1954.

7. Dalman, G. C. and Kao, J. H. K., "Determination of Vacuum Tube Catastrophic and Wear-out Failure Properties from Life-Testing Data," Research Report EE 408, School of Electrical Engineering, Cornell University, Ithaca, New York, Oct. 1958.

8. Pearson, K., Tables of the Incomplete Gamma Function, Cambridge University Press, 1922.

QUALITY CONTROL ENGINEERING IN PRODUCT EVALUATION

E.G.D.Paterson
Bell Telephone Laboratories, Incorporated

This is the seventh and last of a series of papers on Quality Control Engineering sponsored by the Society's Committee on Professional Development. While it is intended to cover one aspect - product evaluation - of the over-all quality control function, the primary purpose of each of the papers is to further the professional standing and recognition of the quality control engineer. To this end, therefore, as well as to the objective of defining the over-all operational framework within which we are to review the particular function of product evaluation, it is first necessary that we agree upon what we mean by the term Quality Control Engineering. That there exists a lack of common understanding in this respect is attested by the fact that, under the title "Total Quality Control", the first paper in this series is directed to a description of the over-all quality control function. (See also Reference I)

One may ask why, especially when the first of this series of papers is directed toward describing specifically the elements of the over-all function, it is proper, as a preamble to a discussion of one of these elements, product evaluation, to make more than passing reference to the whole field. Aside from the fact that we are still in the realm of opinion in this respect, justification is found in the extreme importance of reaching general agreement as to what the job is and what kind of people are required to execute it. Only after such agreement can we determine whether the work is at the professional level, and whether its practitioners are in fact, or should be, qualified engineers in the sense defined by legal licensing requirements. Any material which facilitates this agreement justifies itself on the basis of emphasis, if it is repetitive, and under the guise of clarification, if it is helpfully expatiatory.

HISTORICAL BACKGROUND

It appears to be timely and appropriate under the circumstances, therefore, to review some pertinent historical facts which may not be widely known. As far as the author is aware, the Bell System, and more particularly Bell Telephone Laboratories,* is considered to be the originator of that specialized quality activity which today we refer to as Quality Control in its broader interpretation. It is significant to note that this activity was launched under the then new title "engineering inspection". The breadth of the work envisioned is indicated by the following quotation from a short two page descriptive article which appeared in the August, 1926 edition of Bell Laboratories Record:(2)

- - - - - - - - - - - - - - - -

*Bell Telephone Laboratories was incorporated in 1925, succeeding the engineering department of the Western Electric Company. The latter, as the manufacturing and supply unit of the Bell System, currently has its own Quality Assurance organization with which Bell Laboratories Quality Assurance maintains very close liaison, and from which it obtains inspection results on communication product for the System.

"In the advance of science new fields of activity open, to which new names must be applied. *********. Engineering, the art and science of making useful to man the materials and energies of his surroundings, envelopes the more routine art of inspection; and inspection which is carried out for engineering purpose, becomes "engineering inspection".

****

"In order that engineering inspection shall be of greatest value there must be an art and science of inspection engineering. The pure sciences of mathematics, physics, chemistry and the like must be applied to the problems of inspection in the same spirit and with similar research point-of-view as guide their application to development and design.

****

******"and within our Laboratories inspection engineering, carrying also a responsibility for quality oversight and criticism, is assigned to a separate department. The duties of the department are closely interrelated with those of the development departments and certain technical groups in the Western Electric Company. With proper interpretations they are described as follows:

1. To develop the theory of inspection: putting existing mathematical knowledge into available form for use in laboratory and factory, and developing new principles where existing knowledge is inadequate.

2. To develop methods of stating the quality of various types of apparatus and switchboards, and methods of applying these concepts in every-day work; to develop economic standards of quality which telephone materials should meet.

3. To maintain oversight of the quality of apparatus, supplies and systems currently being furnished for communication service; to make regular reports on the current quality of these materials.

4. To study the quality and performance of telephone plant in service as an aid to further and improved developments; to give special study to instances where unsatisfactory quality is reported and to guide the steps taken to prevent recurrences of such cases."

Now bear in mind that this article was written in the year 1926, some two years after the program had been initiated. Viewed in the light of the circumstances of its time, this new "art and science of inspection engineering" represented a veritable "break-through",- both in the ap-

proach to quality control and in scientific management. If we review carefully the duties outlined, properly interpreting the language as admonished by the article, we recognize as broad a charter for quality control as has yet been claimed by its most ardent advocate. In any event this charter was sufficient to inaugurate a comprehensive quality program which deals with specific phases such as standard setting, scientific inspection, sampling, process control, experimental design, non-conformance handling, quality survey, service performance oversight, complaint investigation, quality evaluation and reporting, inter-phase feedback, and, if the implication is not already sufficiently strong to encompass it, reliability, as has been elsewhere cited.(3) Note also that this program was to be carried out in such manner as best to serve the over-all economy of the Bell System. Although some of us, who have spent our life's work in this broad field, can scarcely look upon it as a "new concept"(4) of the 1950's, we can certainly welcome the strong endorsement implied by its independent development and espousal.

The old adage "handsome is as handsome does" applies significantly in our desire to label our members as professionals. Especially is this true when we have in mind their licensing as Professional Engineers. It is the people, not the job, who are licensed. Let us, therefore, review the type of people who were contemplated as suitable to carry on the work of the "inspection engineer". In a further quotation from the same Laboratories' Record article(2) we find that

> "The fulfilment of these duties requires men of many talents: experts in glass, metals, wood, leather, paper, textiles; engineers experienced in design, manufacture, and operation. Each endeavors to know all that is of engineering interest throughout the history of the design, production, and use of the apparatus or systems in which he specializes. ********, they have no separate laboratories, no individual problems apart from the broad problem of contributing, through cooperation, towards the best telephone instrumentalities which present day science and economics admit".

Recruited to man the organization established to carry on this work were professional people,- mostly from the design, development, and research area, together with a few, who were themselves engineers, from the existing "Inspection" Department of the Western Electric Company. Have in mind that Bell Laboratories has only one general occupational classification for its professional technical personnel. They are all Members or Associate Members of the Technical Staff, whether they be physicists, mathematicians, or chemists,- or whether they be professional engineers, and so licensed, as some are, and some are not. Thus to insure that the caliber of activity carried on in the Inspection Engineering Department would be, as it was calculated to be, at the professional level, professional people were selected at the outset to prosecute it.

It is well to be more specific in order to emphasize the professional character of the work and the workers. Heading the Department with the title "Inspection Engineer" was R.L.Jones, now deceased. Unknown, perhaps, to many of you, although mentioned in the preface of

Dr. Shewhart's text(5) he was previously in charge of Transmission Engineering, author of the Record article referred to, one of the first to be awarded the degree, Doctor of Engineering, by M.I.T., subsequently Vice President of Bell Telephone Laboratories, and, in the author's considered opinion, the person who, more than any other single individual, was responsible for conceiving and initiating development of the broad quality function to which we today address our attention. Among those under his direction, and known to most of you, were such people as George Edwards, previously a Transmission Engineer, widely known for his outstanding leadership in the Quality Control field, Director of Bell Laboratories' program in this area for many years until his retirement, and first President of the American Society for Quality Control; Walter Shewhart, internationally recognized as a statistician, inventor of the control chart, and author of the now classic "Economic Control of Quality of Manufactured Product"; Harold Dodge, engineer, statistician, author, later professor, and widely famous for his sampling and quality rating developments; and Donald Quarles, teacher, engineer, later Vice President of Bell Laboratories, President of the Sandia Corporation, and more recently Deputy Secretary of Defense. The organization was shortly augmented by such well known members of our Society as Doctors P.S.Olmstead, H.G.Romig, and Miss M.N.Torrey.

The character of a job is largely determined by the type of people who carry it on. The formal quality program at Bell Laboratories was of professional caliber from the day it started. It was by design, not accident, that the supporting personnel selected were all professional. The majority were educated and trained in engineering. Indeed, a substantial part of the author's own engineering experience, necessary for qualification for the Professional Engineer's license, was gathered in this Department; buttressed, of course - and this is very important - by the specialized knowledge required for license in any branch of engineering. Thus as a Quality Specialist,* the author qualified for Professional License in electrical engineering. In this connection, it is unlikely that a State will grant a license in quality control engineering unless the applicant demonstrates his ability, by education and experience, to protect the public welfare and safety. Let us keep in mind that, as with a physician's, the engineer's license usually permits the recipient wide latitude in his authority to practice engineering. Its possession carries with it the implication that he is qualified to judge, and ethically so constituted as to confine his activities to, those fields in which he is professionally competent. All things considered, it is the author's opinion that professional recognition of the Quality Control engineer will be advanced most rapidly by having him characterized as a quality control specialist in one of the traditional branches of engineering.

QUALITY CONTROL IS A BROAD ENGINEERING FUNCTION

Referring again to Dr. Jones' article, it is significant to note

---

*In Bell Telephone Laboratories the Quality Control engineer is frequently referred to as a "Quality Specialist" since he directs his efforts primarily to a particular type of product which is referred to as his "specialty".

that there was no mention of either the word "statistics" or the term "quality control". He referred to the activity which he envisaged as a new branch of engineering. What he said was, in effect:

> There is some method - better than the "horse-and-buggy" type of inspection - which should supplement the already recognized efforts of the designer, the producer, and the inspector to evaluate, control, and economically maximize product quality. I conceive it to be a new and separate, but mutually cooperative, branch of engineering. Apply your joint and several professional talents to develop it, organize it, and prosecute it to the end of attaining the highest quality of telephone product economically possible.

Liberally paraphrased though this may be, the author's contemporary discussions with Dr. Jones, and his personal observation of the program's development, convince him that this was essentially the nature of the project which Dr. Jones had in mind.

It was thus as his contribution to this challenge, and as a part - a vital part - of the complete program that Dr. Shewhart had developed the control chart,- a technique finding its place in his over-all statistical control philosophy. Statistics thus provided a scientific tool, a means-to-an-end, a sine qua non, perhaps, but not the complete corpus of the quality control system. Dr. Shewhart himself was well aware of the part that engineers must play in helping to bring to practice the fruits of his own efforts. In developing his theories, he was always solicitous - and considerate - of the reactions of the pragmatic, earthy, sceptical engineers. If contemplation does not convince, experience demonstrates that the phenomenological aspects of all phases of quality control are fundamentally engineering - in import, approach, and treatment.

Widespread and rapid recognition and growth in certain phases of quality control were largely sparked by the "barnstorming" program (1942) of those who may be regarded as the "three musketeers" of statistical quality control - George Edwards, Harold Dodge, and Rupert Gause (the latter then with Army Ordnance, now with Bell Laboratories), undertaken at the request of Army Ordnance and with the whole-hearted support of General Leslie Simon,- a recognized authority on, and ardent proponent of, quality control. However, the breadth of their expository and training activities was severely circumscribed by the exigencies of War (World War II) and the need to concentrate on those particular phases of quality control which could be quickly and widely applied by the Armed Forces in their procurement program so as to provide a maximum of results within a minimum of time. Similar limitations applied, in general, to the programs exemplified by the "eight-day" courses given under the auspices of the War Production Board which did so much to initiate a widespread interest in quality control. Perhaps the necessarily restricted effort and the ever-present semasiological uncertainty associated with the words "quality control" led to the unfortunately narrowed connotation which some now attach to the term "Quality Control". The limb has been mistaken for the whole body!

Bell Laboratories had long since recognized the descriptive shortcoming associated with the term "engineering inspection", had avoided as

inadequately definitive the functional title "quality control", and had coined the term "quality assurance" which, being new and hitherto undefined, permitted the broad connotation with which Laboratories intended to associate it. For those who have not already done so, a reading of the paper "Quality Control in Electronics"(6) should prove to be timely and informative. A review of the literature covering quality control by an author well acquainted with its practice, it points out the intent to attribute to the term a broad quality-assuring function. Similarly, a series of two articles which appeared in "The American Statistician" covers comprehensively and interestingly the more specifically statistical development of Quality Control in the United States over a twenty-five year period.(7)

In his AMA management report paper,(8) Dr. A.V.Feigenbaum points out that

> "----------many businesses over the years have attempted to centralize their company quality responsibilities by organizing a function whose job has been handsomely described as 'responsibility for all factors affecting product quality'. These experiments have had a lifespan of as long as six to nine months - that is, when the job incumbent has the advantage of a strong stomach, a rhinoceros hide, and a well-spent, sober boyhood. Others not similarly endowed did not last the six months. The simple fact of the matter is that the marketing man can best evaluate customers' quality preference, the design engineer can most effectively establish specification quality levels, and the shop supervisor can best concentrate upon the building of quality."

In other words, "each to his last",- most emphatically not "chacun à son gout"!

Secure in the possession of a goodly portion of each of the three advantages enumerated - undisputed as necessary qualifications for any durable quality control engineer - and fortified by the further advantage of over a third of a century in the practice of the function described, perhaps handsomely but not entirely fancifully, as "responsibility for all factors affecting product quality", I feel a strong temptation to vouchsafe agreement (with proper interpretation) with the unidentified author of the quotation. Were one to be permitted the liberty of inserting the article "a" before "responsibility", temerity might give way to smugness. Indeed, having myself used such descriptions as "an over-all review of all factors relating to quality - to the end of getting it and keeping it where it should be" and a "cradle to grave process",(9) I must rise to my own defense. Agreed that the marketing man can best evaluate customers' quality preferences; the Bell System has highly skilled (and professional) groups for just this purpose. Is it not true, nevertheless, that the quality control engineer has a responsibility for knowing this preference, for evaluating it in conjunction with, or at times in terms of, the many specified and measurable quality characteristics, and for directing his efforts towards its attainment? How else would he be able properly to discharge his obligations in setting acceptance criteria, disposing of non-conforming material, reporting on product quality, and initiating any necessary corrective action. I must

disagree categorically with the idea that the design engineer can most effectively, or should, establish the quality levels, if to "quality levels" we attribute the meaning most commonly associated by the quality control engineer. Finally, one need merely pose the question whether the Shop supervisor concentrates on the building of quality with a motivation entirely uninfluenced by his Shop Quality rating.

Now I am sure that Dr. Feigenbaum - whose status as an expert is well secured and who has done so much more than I to spread the broad-concept gospel of quality control - does not find himself in basic disagreement. I indulge this pseudo-facetiousness only as an aid to clarification and to expose the ease with which we presumed experts can maneuver ourselves into an exercise of semantic parry. It is well to remember that the quality control engineer is (at least in the Bell System) authorized to establish acceptance standards, to specify how product compliance with these shall be determined, and to decide whether product which fails to meet these standards in any respect may or may not be shipped. To associate with such authority a concomitment broad-scale responsibility for quality should not over-stress one's credulity.

WHAT IS MEANT BY "QUALITY"?

It is as a part of this very broad concept of quality control that I wish to discuss one of its phases,- product quality evaluation. It will be helpful first to consider briefly just what we mean by the term quality insofar as it relates to our field of activity. Dictionary definitions, of which there are many, fall short of filling our specialized needs. In its ordinary usage a quality is one of a number of attributes or characteristics which, when integrated with others, make an object "what it is" or give it an "over-all quality". It is frequently associated with adjectives such as good, fine, high, etc. which, in the advertising business, seem to appear almost invariably in their superlative forms. Its constituents are often of an abstract nature which defy quantitative measure - style, feel, appearance, workmanship, "the quality of mercy is not strained". In our area of interest, fortunately, we are more often concerned with the utilitarian and more readily measurable qualities which determine such characteristics as operability, efficiency, reliability and life. Nevertheless, such features as the "appearance", "feel", "balance", "fit" and "style" of a telephone handset, for example, are important and much thought, time, and talent have been expended to optimize them. We are not always in a position to neglect the intangibles.

Offhand, one might expect to find the answers to the problem of defining and evaluating quality in the design specification. How else, one might ask, could a product embodying the design intent be brought into being? Certainly, we are warranted in assuming that limits for each of the desired quality characteristics are specified therein or reasonably well implied. And certainly, in the absence of any contradictory information, we should be free to restrict our concern to the particular qualities covered. None know better than we, however, that no two things can be made exactly alike. Hence, no two things have exactly the same quality. Our problem would be simpler, moreover, if it were confined to comparing, and compositely evaluating, product of closely similar characteristics. Actually, we are frequently in need of some

form of quality common denominator among items which appear not to have the remotest relationship one to the other,- at times, it seems, almost as heterogeneous as the alleged ingredients of the male juvenile - "snips and snails and puppy dogs' tails".

As background for our discussion, let us review for a moment the pertinent quality characteristics of a couple of typical products. Let us take first a very simple one,- a sash-weight intended for use as a counter-weight for a double-hung window. To function properly, it must have such shape and dimensions as will insure its freedom to move in the ample slot space provided for it. It must have a relatively high specific gravity in order to reasonably limit its volume. It must have sufficient compression, bending, and impact strength to withstand handling and shipping; tensile strength a little, but not much, greater than the sash cord with which it is to be associated; and a surface configuration and smoothness adequate to avoid handling hazards. It must provide an indefinitely long service-life and it must be cheap. It must have a weight somewhat less than that necessary to balance half the weight and friction effects of the window sash; and its weight must be sufficiently well controlled so that, when associated with its randomly selected balancing partner, no cocking of the sash results. It must be so constituted as to insure freedom from maintenance, since its position in service is relatively inaccessible. Hence its arrangement for attaching the sash cord must be such as to insure the latter's freedom from wear. While it is an end product, note that it functions as part of a system, and that the system requirements have largely dictated its individual properties.

Consider another simple but somewhat more operationally sophisticated item,- the circular disc which constitutes the diaphraghm of a telephone receiver. To fulfill its function as a magnetic-mechanical acoustic transducer, it must have the physical properties which insure a faithful response to the electrical input signal over a frequency range of several thousand cycles per second. (Its acoustical output is, of course, greatly influenced by other parts of the receiver). Its composition and its dimensions are critical. Changes in its characteristics with temperature must be small or suitably compensating. It must resist the fatiguing effect of the millions of vibrations imposed during its expected service life. It must function compatibly with others of the receiver's components which themselves are variable. It must be interchangeable with other diaphragms of its contemporary manufacture. It too, as the sash-weight, must provide a long life in normal service, and it too must be cheap. Unlike the end-product sash-weight, the receiver diaphragm is only one component - a piece-part - of the receiver. Yet its design, and quality considerations are closely associated with its parent receiver, which in turn is part of the telephone set, which is connected through a transmission medium to the complex switching system of the central office whence, through a similar transmission system and another telephone set, it receives, with the aid of the central office power battery, the electrical replica of the acoustic signal emitted by the talking party. (This description impresses one as a technical parody on the humorous - including the humerus - spirtual describing the desiccated osteological articulation of the human).

Since I am not a "quality specialist" in respect to either sash-weights or receiver diaphragms, the foregoing analysis of their respec-

tively necessary quality characteristics may be presumed to be incomplete. Yet it will demonstrate the many design (and quality) considerations which are involved in even the apparently simple items; and it will demonstrate, in particular, the type of thinking which is preliminary to the exercise of any one of the quality control activities. To lay out and prosecute properly the quality control program, it is vital that the quality control engineer have a thorough technical knowledge concerning the particular product involved; he must be familiar with and understand not only its design but its design intent as well. If he does not already possess such specific knowledge, he must acquire it; and to acquire it he will usually need a sound, basic, technical training. This is illustrated in the following excerpt from an article describing Laboratories' Nike Quality Assurance program:(10)

> "The Nike missile is a complex electromechanical device and, while our quality assurance engineers have had wide experience with such devices, the missile introduced to them new physical interrelationships and, in particular, new aerodynamic and ambient environments. It is axiomatic with us that our quality specialists must be familiar with the design, function, manufacture, inspection, and use of their specialties."

QUALITY EVALUATION

At the outset, let me make it clear that in the material to follow, I have attempted primarily to discuss the type of thinking and rationalization processes involved in product evaluation. I have purposely avoided either prescribing or delineating any specific operational plan, and I do not mean to imply that any operational procedures which I may suggest are necessarily an actual part, or are handled precisely in the manner indicated, in any existing evaluation plan. There is a relative dearth of literature on this subject and, until the results of more widespread experience in this area are forthcoming, it would be unwarrantably presumptuous to be overly didactic.

The character of the comments and questions from several of my colleagues after their reading of my first draft of this paper convinced me that I was not warranted in assuming that all of us attach one and the same meaning to the term "quality evaluation". My somewhat indiscriminate use of the terms "quality evaluation" and "quality rating" had led to confusion. Toward a better common understanding therefore, let us review some of the factors involved in any attempt to "evaluate" quality. Basically, we, as quality control engineers, are concerned with that interpretation of the word "value" which represents a measure of the degree to which a product satisfies or fulfills an economic need or want. This economic need or want cannot ordinarily be defined by a single attribute of the product. One is usually interested in how a pair of shoes "look" and "feel" as well as in how they wear. Moreover one's concern with respect to these characteristics may vary from person to person, from time to time, and with circumstances. The farmer in the field values the working brogans above dancing pumps. We all recall the oppressed equestrian willing to exchange his kingdom for a horse - an excellent illustration of the value of time and place utility. Not only is "value" a relative term, but most of those of its constituents which are sub-

jective in nature do not lend themselves to physical measurements. The "value" of a product is thus relative, subjective, a function of time and circumstance, and an assessment which cannot be expressed as a single quantity. Despite the misgivings which these limitations engender, the job of quality evaluation is not altogether hopeless from the standpoint of the quality control engineer.

For purposes of our discussion, we can divide the problem of quality evaluation into two categories: (1) the assessment of those features which determine the customer's preference for, or the economic superiority of, product A over a similar product B; and (2) the assessment of the variations in relative quality of a product of essentially the same design, from unit to unit, or lot to lot, or time to time. In either case, we may be interested in an individual quality characteristic, or in the over-all quality which results from the combined effects of all of the characteristics. In the first category, we may compare Fords with Plymouths and Chevrolets, or the 1958 model with the 1959 model. In the second category, we may compare 1959 Fords made today with those made yesterday. In the first case, we are concerned not only with variations from one Ford to the next but, more particularly, with basic differences between makes or models. In the second case, we effectively disregard the basic differences. While the quality control engineer is interested in both categories of evaluation, prior decisions on the part of the market analysts and research and development engineers will have essentially fixed the basic designs, and thus the broad quality differences envisioned in the first category. Quality evaluation in this more fundamental area involves so many non-technical variables that, so far as I am aware, no broadly applicable, quasi-quantitative procedures have been developed for the purpose. Quality evaluation of the type included in the second category, however, is a procedure which more easily lends itself to formalization, is the process commonly understood by the term "quality rating", and is normally part of the assigned responsibility of the quality control engineer. It is this latter type of quality evaluation which I will discuss in some detail.

QUALITY RATING

The over-all quality of the sash-weight, or the receiver, is a complex composite of many separate contributory quality characteristics, each of which is itself a variable,- not usually an independent variable_ and each of which has a significance depending largely on the assembly or system of which the individual product is a part. Thus the measure of the composite quality of an individual unit of product may not only have any one of a potentially infinite number of values, but its relative importance may also depend upon the particular qualities of a large number of separate and dissimilar products with which it is to be associated. The designer must recognize these relationships and potentialities and specify suitable requirements for the individual unit to assure satisfactory performance. For quality rating purposes, the quality control engineer must not only recognize these relationships, but must attempt to take into account quantitatively their relative importance to the over-all quality of the product. He must do this in a manner which reconciles the intent of the designer, the problems of the producer, and the desires of the consumer, and in a manner which recognizes the exigencies of, and accords full importance to, economic costs.

The over-all product quality, and its constituent characteristics, are usually not specified in terms associated directly with the ultimate consumer's expressed desires. The consumer wants a telephone which reproduces the voice "clearly" and "naturally". He want an automobile which will accelerate rapidly, provide so many miles per gallon of fuel, ride comfortably, and have a pleasing appearance. The designer specifies limiting values for the volume response of the receiver over an appropriate range of frequency under prescribed conditions of input. For the automobile, he specifies torque, power, etc. (these in turn, are usually insured by other engine design requirements) to control engine performance; weight distribution, spring characteristics, etc. to control riding qualities; and dimensions, materials and processes to control shape, style, and finish. It is customary, and quite proper, for the quality control engineer to assume initially that the basic quality considerations are appropriately defined, circumscribed, or implied in the design specifications. He may subsequently find this assumption to be unwarranted in some instances - that the specifications are inaccurate or, more frequently, inadequate in some respect. Review of the design specifications for just this purpose is a part of another phase (quality survey) of quality control. In the event of such error or inadequacy, the remedy is obvious, though not always simple. At any rate, the point to be kept in mind by the quality control engineer is that, while the specified characteristics normally provide the basic parameters by which quality is to be evaluated, the significance of the quality characteristics should be measured in terms of their ultimate effect on service performance in which they fulfill a "means-to-an-end" function. This, of course, involves a problem of weighting.

Because the complexity of quality rating increases rapidly with the increase in the number of qualities to be measured and combined, there is a strong incentive to minimize the number of characteristics to be treated as independently significant for quality rating purposes. Whatever can be done to select one characteristic as representative of several others with which it is sufficiently well correlated - or to eliminate entirely attention to individual or over-all qualities which can properly be assumed to be fixed for rating purposes - will greatly simplify the evaluation process. This applies particularly to piece-parts or components which function "passively", (this term being construed for present purposes in the lay sense to include structural parts such as framing, screws, insulators, etc., but not circuit components so characterized in the more specific technical connotation). The telephone receiver diaphraghm, for example, is characterized by material and dimensions. The material, in turn, is dependent upon raw constituents and processing,- the latter involving mixing, metallurgical and mechanical processing, and heat treatment. If we can assure that the material, in the comprehensive sense indicated, is uniform and correct, then the magnetic-acoustic properties of the diaphraghm itself may be related exclusively to dimensions. If these, in turn, can be sufficiently well controlled, we are warranted in omitting individual quality evaluation of the diaphraghm and in confining attention to the quality of the receiver of which it forms a part. It requires little imagination to visualize the practical impossibility of rating the quality of each characteristic of each part of each product of the total output of a large enterprise. The first problem, then, is to determine to what detail it is necessary, and practically reasonable, to go in order to insure an adequate and dependable quality rating pro-

gram. On the one hand, we are constrained by the costs involved in meticulous attention to details; on the other, we are urged by the specter of the potentialities of oversight. The undetected omission of the specified heat-treatment of the clamping screws in the spring pile-up of a relay, and their subsequent "head-popping" as a result of the hygroscopic expansion of the pile-up separators under humidity in service, completely vitiates the carefully compiled quality evaluation of the relay based upon its dynamic end-tests in the completed state. This problem as to how far to go in the selection of pertinent quality information for evaluation purposes calls for the highest degree of engineering judgment,- a judgment which results from the quality control engineer's "endeavor to know all that is of engineering interest throughout the history of the design, production, and use of the apparatus or systems in which he specializes".

From what has been said, it follows that one, if not necessarily the only, technique of rating the quality of an individual unit of product is to measure the actual values of the several carefully selected characteristics, weight each of these values by a factor which takes into account both its variation with respect to its own norm or standard and its significance with respect to the other characteristics, and finally to combine the weighted contributions, thus obtained, into an over-all value - a number or some form of quality index. The resulting quality index for the one unit can be combined with the other (N-1) quality indexes in the group of N units to obtain an average quality index for the whole group.

As has been indicated, one of the primary purposes of quality rating is to enable us to compare current product quality with what we believe it should be,- to enable us to determine whether it is "good", "satisfactory", at the "desired" or "aimed at level" (to use commonly employed terminology), or how, by how much, or how seriously it deviates therefrom. We require, hence, some standard of reference with which we can compare current quality. The standard may relate to an individual characteristic, to over-all quality of an item of product, to groups of similar products, or to aggregates of dissimilar products. As has been pointed out by Dr. Shewhart,(5) the standard should be expressed in terms which are meaningful, consistent with custom, enforceable, and sufficiently broad to allow rational interpretation under different circumstances. It should represent a value which results when the combined economy of production and use is as nearly optimized as our knowledge permits. A little reflection will indicate that, since circumstances of production and use are constantly changing, and since our knowledge of all pertinent economic factors is by no means precise, our standards will be subject to continual revision. The setting of economic standards of quality is decidedly not an exact science.

Having established the standard of quality for a particular product, we can then proceed to compare current product with this standard. We might consider each of the quality characteristics in a positive manner by assigning to it "merits", increasing the number of "merits" as the value of the characteristic gets closer to the desired or standard value. Or we can on the other hand - and this method has had a long and successful history - assign an increasing number of "demerits" as excursions from the desired value become larger. In either case, the number of "merits" or "demerits" to be assigned will be affected not only by the

degree of departure from its own "standard" but also by the significance of the particular characteristic with respect to the over-all quality of the product.(11)

We can simplify our evaluation process tremendously if we can justify ourselves in making what, in some cases, may be gross approximations,- if we can assume, for instance, that all units meeting the specified requirements are equivalent in quality. In this case, we can base our quality measure for a group of similar units on the proportion of units in the group which have one or more individual characteristics with values outside of the specification; in turn treating those units which are outside of specification either (1) as of equivalently poor quality or (2) as differing in degree of poor quality on the basis of the extent of their departure from specification. This procedure is not properly applicable - or at least by itself not adequate - in those many cases where the specification limits are not, by themselves, sufficient to insure an acceptable level of quality. The average value of cross-talk on loading coils, for example, must be well within the specified limits (which apply to the individual coil) to insure the required quality of transmission of the lines in which a number of the coils are to be used.

One rating method expresses the quality significance contributed by a particular characteristic (or product) in terms of its excursions from a standard value, the excursions being measured in multiples (t) of the standard deviation (σ) of the measured characteristic (or product quality). This has many practical advantages: it is simple to handle statistically, is acceptable to the Shop people to whose product it is applied, and fulfills its major purpose - control of the product at the specified quality level. (This method of measuring the significance of the departures from the standard in terms of multiples of the standard deviation should not be confused with t_σ as associated with confidence limits). Another technique employs an index expressed in terms of the ratio of actual demerits found in the product as compared with the number of demerits expected when the characteristic or product in question is at the standard level.

While much of the discussion has been, and must continue to be, concerned with the principles of quality weighting - since weighting is involved in the process of combining the effects of the various quality characteristics - we must never lose sight of the importance of the individual characteristic. In some circumstances, obviously in the case where quality rating is directly related to the purpose of process control, major attention will focus on a particular characteristic. New designs, redesigns of current products, field complaints, or excessive nonconformance often prompt such attention. The quality control engineer must, of course, be familiar with the status of each of the important characteristics, in any event, lest he be misled by over-all composited rates. This is a case where it is necessary to see the trees as well as the forest.

OVER-ALL PRODUCT QUALITY SHOULD BE ECONOMICALLY MAXIMIZED

The problem of product quality rating would be complex even were it to be confined to the rating of individual product units and groups or "lots" of the same or very similar units. In the usual case, however, there must be provided, in addition, some means of evaluating and por-

traying the composite or over-all quality of heterogeneous products. For example, in the communication business, we are concerned with the quality of thermistors and clay conduit, aromatic spirits of ammonia and klystrons, telephone poles and central office systems, nitrogen gas and diesel engines, etc., etc., - not in these particular combinations, of course, but individually and, together with many others, collectively - today, next month, and from year to year. Management needs some measurement symbol - a chart, an index, a number - which will indicate, relative to some standard of goodness or economic soundness, the quality of the aggregate output in a given production period; that is, some symbol which will best represent the combined effect of the individual variations in the quality of the several products.

The goal of each product is to provide customer satisfaction and, for most of the products with which we are concerned, customer satisfaction is measured in terms of service. The term "service" is used here in the broad sense to include those factors, such as appearance, to which customer reactions are esthetically motivated. Whether it be a telephone or a washing machine, the customer looks for satisfactory performance over a reasonably long, trouble-free,period. He is equally interested in another factor - cost. It is a fair generalization to state that we can almost always improve service at the expense of added cost. In the final analysis, the customer, and thus the Supplier's management and his quality control engineer, are interested in maximizing the service per dollar ratio. Under these circumstances, in the face of a lack of any absolute measure of quality, we are easily led to adopt the dollar value of the product as the common factor which will enable us to combine the otherwise often remotely related product qualities. This presupposes, of course, that the quality standards for the individual products will have themselves been set at levels which represent optimized quality per cost; where cost includes not only the research, development production, and marketing expense but the expense associated with installation, operation, maintenance, depreciation and obsolescence as well. In theory then,- and in practice as closely as we are able, and are justified, to approach the theoretical - we establish for each product a standard of quality which represents an optimum economic level. In the Bell System, we compute an index of quality for each product by comparing the observed value with its standard. We derive an over-all product quality index by weighting the contribution of each product's quality index by the value of the product, averaging, and comparing the over-all average with an over-all standard.

I hasten to point out that I am well aware that some may take issue with the propriety of this combination on the basis of dollar value. A fuse which carries current well beyond its designed limits and thus fails to protect expensive equipment, or a weak lineman's safety strap which introduces a personal accident hazard will, on a dollar weighting basis, have little effect on the over-all quality picture which covers large quantities and varieties of expensive product. Safeguards against such irregularities lie in anticipating their possible occurrence, and in the employment of acceptance criteria and procedures which will, to the extent practicable, obviate their presence in accepted product. The technique of quality evaluation - and its consequences - should be reflected in all facets of the quality control job,- in the prescription of acceptance sampling plans, the disposition of non-conforming product, the re-

vision of a manufacturing process, etc., etc.

Reference has been made herein to quality standards which represent desired values for individual characteristics and for over-all quality, and to those which represent desired levels for aggregates of various products. Since each of these standards bears a close relationship to the design requirements, and to the production facilities available for implementing them, and since the design requirements and facilities change from time to time, the quality standards are not static. They represent a level of quality, in terms of the existing design, which it is desirable to maintain under existing circumstances. It is unsafe, therefore,- actually it is often erroneous - to compare a product's quality rate in one period with its rate in a different period and assume that one gets thereby a reliable measure of the change in the product's "intrinsic excellence". What was considered to be a "good" automobile twenty years ago is quite different from its "good" counterpart of today. Also, the over-all quality standards for combinations of more than one type of product must vary from time to time as the individual product unit costs and output quantities vary. Aside from these considerations, it must be appreciated that the pathway to quality rating is by no means straightforward or clearly marked. Each step is subject to error: error in measuring, error in assaying the capabilities of the production process, error in judging the relative importance of quality characteristics and their deviations, and so on. Indeed, the very uncertainty of the quality rating course, and the necessity for the continual correction of our movement, persuades us to use all the available talents of the professional guide, that the progress of our never-ending journey may keep us as near as possible to our intended destination.

It is doubtful that there exists any single rating system which, in whole or part, represents the best technique for all circumstances. Any system which sets out to depict over-all product quality as a point or a single value is subject to the shortcoming inherent in any average, i.e. the good hides the bad or vice versa. The various techniques, herein referred to briefly, have at least the merit of successful use; they have all contributed to the control of quality at a suitable level. As a matter of fact, it is not the intent of this paper to describe or compare any specific methods employed in quality rating technology. Deliberately, they are discussed sparingly, sketchily, and only so far as clarity demands. The primary purpose of this paper is rather to examine the principles underlying product quality evaluation; to indicate the technological, mathematical, and statistical complexity* inherent in the problem

* I am indebted to my colleague, Dr. R.B.Murphy for the following gustatory metaphor which I quote with his permission: "It seems to me the most profound complexity is trying to preserve a rational relation between the real world and the quality evaluation technique. This problem is an odd gumbo of technical, economic, and psychological vegetables. The job is nevertheless necessary, and my feeling is that to determine whether the ingredients have been properly combined and seasoned can be done only by continuous tasting - to see whether in the long run we have created a "fair impression" that has led to "fair results", and this is a matter which must depend not only on our knowledge of the technical situations which existed and contributed to the evaluation but the manner in which these data were compressed in the evaluation and the result of such compression on the reading public and the people whom the reading public affect in turn."(12)

of designing and implementing any mature quality rating program; and thus to demonstrate the professional character of the people equipped to carry on such a program properly.

Scarcely any reference at all has been made to the primarily statistical aspects of the rating technique. When we reflect upon the fact that each of the quality characteristics with which we deal is a variable, that its value is uncertain, that the combinatorial techniques can be mathematically intricate, and that we wish to associate with the ultimate quality rates a degree of confidence which will have operational meaning, we can appreciate the statistical skills which are required. Indeed the technological difficulties inherent in sound product quality rating in both the engineering and statistical area - are so profound as to account for the source of my impression that - of all phases of quality control - the formalized process of quality evaluation is farthest from its ultimate maturity. When the chips are down, the best single picture of the over-all quality of any product derives from the judgment, perforce qualitative, of the individual quality control engineer within whose responsibility the particular product lies. His judgment is supported by the quantitative information afforded by formalized rates, but it must be tempered by information from such sources as quality surveys and engineering complaints, which presently find little application in the formalized rating process. So it is, then, that an over-all "Quality Report" should include not only the quantitative information depicting individual characteristics, products, and over-all output, but also supplementary qualitative and interpretive statements which are based on the quality control engineer's knowledge derived from sources which are not, nor likely to be, amenable to rigorous statistical treatment.

ENGINEERING vs. STATISTICS

It should be clear from what has been said that product quality rating presupposes the exercise of a high degree of expertise in two fields, broadly characterized as engineering and statistics. The necessary combined knowledge may be found in the one individual. But especially as the quality control job takes on more maturity, as it demands original thinking of a professional character rather than mere utilization of an already routinized procedure, there is likely to appear a specialization in each of the two broad fields. This does not mean that the one specialist will be unfamiliar with the other's primary forte. Actually, to insure the necessary degree of cooperative effort, there must exist a substantial over-lapping in proficiency as well as in sympathy.* Even within the engineering field of quality control, as the employment advertisements bear witness, a degree of specialization is advantageous if not actually necessary. The mechanical engineer, per se, would be ex-

- - - - - - - - - - - -

* This necessary close cooperative relationship is excellently espoused and illustrated in an article "The Relation Between The Engineer And The Statistician" by Joseph Mandelson - Industrial Quality Control Vol. XIII, No. 11 - May, 1957. Incidentally, on the occasion of an address before the Army Chemical Center, Md. several years ago, I was surprised and pleased - to learn that their underlying philosophy, as well as many of the details of their organizational set-up and practice of Quality Assurance closely paralleled those of Bell Telephone Laboratories.

pected to find himself more at home with the quality engineering aspects of automotive gear than with those of the guidance section of a missile. At least at Bell Laboratories, we found it necessary, as anticipated in the article by Dr. Jones, to provide ourselves with "Quality Specialists" in the sciences and engineering as well as with specialists in mathematical statistics.

If "quality is everybody's business", as it should be, there must be some distinguishing characteristic attributable uniquely to "quality control" to justify its continued existence. This characteristic, the author believes, is to be found in its "over-all", detached, objective approach to quality, to the end that the integrated individual quality responsibilities of design, production, and use are, as closely as possible, economically optimized. Thus quality control in its most widely used industrial applications involves (1) the ability to recognize, measure, and evaluate the quality aspects of design, production, and use, i.e., engineering, and (2) the ability to recognize, measure and evaluate the significance of relationships, i.e., statistics. Quality Control was born as a result of the marriage of these two. It is highly probable that a technological expert other than an engineer might properly replace the latter. The statistician - or his equivalently informed counterpart- must always be available despite the fact that, as a professional, and as a statistician, he may not be qualified for engineering license.

Significant contributions to the field of quality control have been and will continue to be made by professionals of many types. The Society must be careful, in its endeavor to advance its recognition as professional, not to jeopardize the status of those members of the highest competence who might not qualify specifically as professional engineers. In particular, it must not segregate those respected members - the professors, who, in any event, do not need additional testimonial to their professionalism. I cannot forget that my first consciousness of the type of thinking which underlay the development of what is now known as quality control occurred many years ago, when, as a student in electrical engineering, I was exposed to a course in "Precision of Measurements" taught by a professor of physics!

SUMMARY

The underlying purpose of this series of papers is to disclose the professional aspects of the activity referred to as Quality Control. There exists a wide variety of concepts as to the scope of this activity- its responsibilities, its modus operandi - and there exists, in consequence, some uncertainty as to the necessary qualifications for its practitioners. In these circumstances, I have felt it apropos and timely to preface my text on product evaluation with some history concerning the introduction of a quality control activity which has been carried on as an independent function by a group of professionals continuously over a period of thirty-five years. The manner and substance of the presentation should make it apparent that my opinions and convictions are deeply and indelibly colored by my experience in one industry - the Bell System. In fact, I have used the term "Quality Control" herein in the same sense as "Quality Assurance" as practiced by Bell Telephone Laboratories.

It was not necessary to "sell" quality control to Bell Laboratories'

management. Management conceived the activity, directed that it be developed and presecuted, assigned the professionally qualified people to do the work, gave them status as full partners in an over-all development and research organization, and provided them the authority necessary to the success of the operation. From such a propitious genesis - nurtured by the whole Bell System's continuing keen interest in quality, from raw material to ultimate communication service - one might anticipate a program of lasting fertility. Whatever success it has attained is due in large measure to the steadfast support and encouragement which it has received from management.

It is appreciated that the particular organizational set-up of the Bell System is especially favorable to an adequate over-all quality control program. The objective of the entire System is to provide the best communication service consistent with its cost. This is the guiding directive of the parent American Telephone and Telegraph Company. This is the purpose of the research and development activities of Bell Laboratories, of the purchasing, production, installation and repair activities of the Western Electric Company, and of the several telephone operating companies which require presently almost two billion dollars' worth of additional communication product annually to implement this purpose. Despite the designer's natural pride of creativeness, the producer's pride of his skill in manufacture, and the operator's pride in his utilization know-how, each is willing to subordinate his individual prerogatives in the interest of maximizing over-all economic quality. This close community of endeavor provides an unusually favorable atmosphere for a completely integrated quality control program. I recognize that the practice of quality control is not always carried on under these congenial circumstances, which so greatly simplify the implementation of quality controlling decisions. This deficiency does not impeach the validity of the principles of quality control engineering; it merely makes their effective application more difficult.

Quality control represents an over-all program comprising a number of closely integrated functions. It may be compartmentized only for the purposes of analysis and ease of description. One of the functions - one which courses continuously through the whole control system - is product quality evaluation,- more specifically, what has been referred to herein as quality rating. Successful quality rating presupposes a mature knowledge of the quality aspects of design, production, and use, and the exercise of the professional talents commonly associated with the terms "engineering" and "statistics". Proper execution of the other functions of quality control requires knowledge of the same level and nature. It is my belief - and, at least in the Bell System, long experience has demonstrated it to be true - that specialization should, except perhaps in the statistical area, be on a product basis rather than on function.

By citing the fact that Quality Control was conceived as a new branch of engineering, was staffed by an organization predominant in engineers, and requires for its functioning the type of knowledge commonly referred to as engineering, I do not intend to imply a secondary importance to the statistical aspects of the job. Quite the contrary! None is better aware than I that this new "branch of engineering" originated from melding these two fields of art and science. In my long ex-

perience, I have yet to see a difficult problem in the application of quality control properly solved without the contributions of both the primarily engineer and the primarily statistician. Indeed the success of the whole quality control program depends upon the ability of the quality control engineer to search out and amalgamate knowledge and help from each of the sources whence ultimate quality traces its growth.

A society is a professional society if its members are in fact professional,- whether or not they be legally licensed as such. If the quality control program is carried on, as it should be, at the professional level, then its practitioners must, ipso facto, be professionals. On the other hand, a job's stature is determined largely by the caliber of those assigned to it. There arises, thus, an enigma akin to that posed by the question concerning the antecedant priority of the hen and the egg,- and likely of about equivalent significance. The point is that the Society should so define the technique of quality control as to necessitate full professional qualifications for its exercise. "By their fruits ye shall know them".

While attention has been drawn frequently herein to those personal attributes which are primarily associated with, or are of importance to, the more purely technical aspects of the quality control job, scarcely even allusion has been made to those which fall under the broad classification of ethics,- those which determine what the layman associates with a person's "character". It is difficult to visualize a legitimate pursuit which will not benefit from any of the ethically desirable traits in its personnel. Quality control is decidedly no exception. Since in the normal course of his duties, however, the quality control engineer must call attention to, if not actually stigmatize, the results of others' shortcomings, and since he must pointedly, although constructively "needle", he must be completely and honestly objective; he must act, unvaryingly and convincingly, in a manner which attests to his unquestioned integrity. Thus motivated, his honest errors will be excused, his pestiferousness tolerated, and he will have the key to the warehouse of cooperation from which he must draw in such abundance. The aforementioned "well spent, sober boyhood" is certainly conducive to such attainment, though it may not actually constitute what the lawyers term a "condition precedent".

ACKNOWLEDGMENTS

I wish to express my appreciation for the constructive comments and suggestions of my colleagues, C.E.Fisher, G.R.Gause, R.C.Koernig, R.B.Murphy and Miss M.N.Torrey concerning the first draft of this paper. These are the people who actually see to it that the difficult job of quality rating is properly carried on for the Bell System.

REFERENCES

(1) "The Professional Work of the Quality Control Engineer" - Dr. A. V. Feigenbaum
Industrial Quality Control - Vol. 14, No. 8, Feb. 1958

"What is Quality Control" - C. C. Craig
Industrial Quality Control - Vol. 15, No. 1, July, 1958

"The Quality Control Budget" - Frank H. Squires
Industrial Quality Control - Vol. 15, No. 6, Dec. 1958
(The author of the present paper on Product Evaluation does not subscribe to the narrow definition contained in this reference that "quality control is essentially a production technique aimed at preventing the manufacture of defective parts".)

(2) "The Viewpoint of Inspection Engineering" - R.L.Jones -
Bell Laboratories Record - Vol. 2 - No. 6 - August, 1926

(3) "Observations on Quality Assurance and Reliability" -
E.G.D.Paterson
Tooling and Production - Vol. XXII - No. 12 - March, 1957

(4) "Quality Control, Principles, Practices and Administration" -
A. V. Feigenbaum
McGraw-Hill, 1951

(5) "Economic Control of Quality of Manufactured Product" -
W. A. Shewhart
D.Van Nostrand Co., Inc. 1931

(6) "Quality Control in Electronics" - Mary N. Torrey -
Proceedings of the I.R.E. - Vol. 44 - No. 11, November, 1956

(7) "The Development of Statistical Quality Control in the United States"
S. B. Littauer
The American Statistician, December, 1950, February, 1951

(8) "Total Quality Control - A New Concept of Integrated Quality Control Engineering" - A. V. Feigenbaum
AMA Management Report, Number 9, 1958

(9) "An Over-All Quality Assurance Plan" - E.G.D.Paterson -
Industrial Quality Control - Vol. 12, May, 1956

(10) "Nike Quality Assurance" - E.G.D.Paterson
Ordnance - Vol. 41, November-December, 1956

(11) "A Check Inspection and Demerit Rating Plan" - H. F. Dodge and Mary N. Torrey -
Industrial Quality Control - Vol. 13, July, 1956

(12) Private communication.

TOTAL QUALITY CONTROL

A. V. Feigenbaum
Manager - Manufacturing Operations
and Quality Control Service
General Electric Company

The mid-twentieth century American economy has been described as one of "Rising Expectations" on the part of its public. Nowhere is this more clearly demonstrated than in terms of the quality demanded by the American consumer in the products he buys. Year by year, and whether housewife or large corporation, the consumer has been asking for progressively higher levels of product quality. Inevitably in our free economy, he has been receiving what he demands.

Omitting inflationary factors (by considering real prices relative to real wages) technical evidence proves that our refrigerators, our turbine-generators, our steam irons and our fry pans have unquestionably provided us in each succeeding year with better, more consistent, day-in-day-out value per price dollar -- which is, after all, a practical definition of quality.

Thus, the 1959 era of competition in the American market place is fully as much an era of quality competition as it is one of price competition. Recognition of this fact -- and dealing with it adequately in business planning and action -- is, I propose to you, one of the great management challenges of our day.

Consider this fact: marketing forecasts predict that customer demands toward higher and higher quality will intensify, rather than diminish, in the next few years. Certainly, the electrical industry typifies this trend. Quality levels for motors and generators tighten appreciably year after year. Consumers are progressively more critical in their examination of the finish of appliances and in their judgment of the fidelity of radio and television sets. Even for those military products on which human safety already depends -- jet engines, airborne electronics, naval ordnance -- quality requirements are being upgraded.

A key result of this increased customer demand for higher quality products has been a rapid and dramatic obsolescence of in-plant practices and techniques for controlling the quality of the products being manufactured.

For example, the machined part for air conditioner compressors that could once be checked with precision micrometers must now be carefully measured with an air gage. Spring material that could once be visually accepted if it were "reddish brown and shiny" must now be carefully analyzed both chemically and physically to assure that it is beryllium copper instead of phosphor bronze.

Performance testing requirements call for more and more complex testing equipment. Intangibles such as dust in the air and humidity have become extremely tangible quality problems, and the objective of eleborate safeguards.

Automation, in which rapid quality evaluation is a pivotal point, has magnified the need for mechanization of inspection and test equipment -- much of which is now in the hand tool stage.

As a major consequence of this in-plant obsolescence, quality costs (inspection, test, laboratory checks, scraps rework, customer complaints and others) have -- largely without fanfare -- crept up to become a multi-million dollar item.

The total of these quality costs for many American businesses compares today with their total direct labor dollars, total distribution dollars, and total material purchase dollars. Although there are no formal nation-wide studies on the subject, evidence suggests strongly that many businesses have quality cost expenditures representing 8%, 10%, 12% and much more as a percentage of their net sales billed. In some cases, this means 80%, 90% and 100% and even more as related to their total direct labor dollars.

So, collectively, these three current trends: higher quality demands by customers, the obsolescence of company quality practices, and the resulting high quality costs, confront the American businessman with a twin-edged quality problem:

FIRST: Much improvement must be effected in the quality of many products and in many quality practices.

SECOND: These quality results must be accomplished simultaneously with substantial reductions in over-all quality costs.

Business management from Coast to Coast is intensively and aggressively turning to better ways and means for meeting this double-edged problem. Let's review the direction of this effort.

Management has started by accepting the principle that the task of controlling product quality -- that is, the task of quality control -- is far more than a matter of what we have historically known as inspection work. For the quality of a product -- an air conditioning unit, for example -- has been affected at many more stages of the production cycle than those susceptible to inspection activity.

The total weight of quality is actually created in seven sequential stages.

Quality is influenced by the Marketing function, which evaluates the level of air conditioning quality which customers want and for which they are willing to pay.

Quality is influenced, to a major degree, by Engineering -- which must reduce marketing's evaluation to exact drawing specifications.

Quality of our air conditioning unit is influenced by purchasing, which must choose, contract with, and retain vendors for parts and materials.

It is influenced by Manufacturing Engineering, in its selection of the jigs, tools and processes which will vitally effect quality throughout production.

Manufacturing supervision and shop operators have had a major quality influence during parts-making, sub-assembly, and final assembly.

It goes almost without saying that the job of mechanical inspection and functional test is necessary to check the product's conformance to specifications.

And finally, Shipping has its influence on quality -- an influence seen in the calibre of packaging and the excellence of transportation.

So, the determination of quality -- and, consequently, of quality cost -- takes place throughout all stages of the production cycle. The twin objectives of better product quality and lower quality cost cannot be achieved by concentrating upon any one phase of the cycle alone -- inspection, design engineering, reject trouble-shooting, operator education, or statistical analysis -- important as each phase is in its own right.

Doing the job adequately requires instead,a program of total quality control.

Therefore, I propose to you that total quality control -- through quality control engineering -- not only is the function which can do the quality control job needed by American business today, bu t it is also the activity that is geared to the quality job that will face us tomorrow. As an inevitable corollary, it is the quality control concept that is most acceptable to American business: because it is an integrated whole; because it operates at lower costs; because it contributes in major degree to business profitability; and because it comes to grips with product quality and quality cost problems as top management sees these problems.

This broad approach welds the new technology of quality control into an organization structure that has sufficient strength and is broad enough to come to grips successfully with the three major 1959 quality problems that modern competitive business must face and solve; the upward customer pressure on quality levels; the resulting rapid obsolescence of quality practices and the extremely high level of quality costs.

Although, delivery, promotional stunts, and other factors may sell a product the first time, it is usually quality that keeps the product sold and that keeps the customer coming back a second, third and fifteenth time. Quality cost -- perhaps 8% or much more to cost of sales -- is one of the major elements of product cost that must be made right to permit the setting of the right price to the customer.

Helping to assure this right quality at the right quality cost is the way in which the new total quality control must, in the years ahead, serve American companies and the American economy's progressively lower real costs.

AIR FORCE QUALITY CONTROL ACTIVITY IN RELIABILITY

O. C. Griffith, Colonel, USAF
Quality Control, Hq AMC

Introduction

Gentlemen: It is indeed a pleasure to return again to an assignment in Quality Control, and to renew many old acquaintances with the members of the American Society for Quality Control. I have been impressed with the amount of change in Quality Control in the few years that I was not active in this field. It is an excellent indication that Quality Control is in fact a dynamic and growing field in the industrial arts. Today I would like to discuss with you the work we in Air Force Quality Control have underway as our part of the logistic team effort so necessary if we are to increase the reliability of today's weapon systems.

It is almost trite to say that we, just as many other organizations in the logistics business, have been continually increasing our effort in reliability in the recent years. Before delving into the specifics of this activity, I'd like to take the time to review briefly the basic premises upon which we are orienting our work in reliability. These premises are in no way new or novel, but there does seem to be sufficient divergency of thinking in some areas to warrant spelling out the framework within which we in Air Force Quality Control are operating.

The first major premise is that reliability in a technical sense represents a major problem, cutting all parts of the logistics cycle. These technical areas will require a continued major effort for a considerable amount of time.

The second major premise is that reliability represents a major challenge to management. It is widely recognized, for example, that management cannot literally divorce the reliability engineering function entirely from that organization charged with overall engineering responsibility and authority. The same facts of life are, of course, recognized to be true for the other logistic functions such as procurement and quality control. The reliability implications inherent in so many aspects of logistics activity is a serious challenge to the accomplishment of the teamwork and coordination necessary to achieve reliability within our available resources.

General Reliability Activity

Turning now to areas of more specific activity, I would like to discuss briefly a number of major actions we presently have underway.

I will divide these into two groups. The first group is primarily managerial in nature, while the second is primarily technical or scientific.

Internally within the Quality Control Office at Hq AMC, we have established a Reliability Project Group. This group has been active

for some time; it is responsible for monitoring all reliability activity within the office, and for planning and developing quality control concepts, policies, and programs concerned with reliability. Establishment of this group has greatly facilitated our effort to increase the emphasis in reliability.

This group is also working very closely with other AMC and ARDC organizations in order to achieve the coordination and teamwork so essential to many of our reliability problems. Major examples of such specific activity are efforts toward development of completely integrated reliability programs within AMC, team review of contractor reliability programs from a management viewpoint, and collaboration with ARDC elements in the development of management type reliability specifications for use in weapon system and equipment contracts. We consider the development of such specifications one of the fundamental steps in the Air Force reliability effort. I will discuss such specifications in greater detail later in this paper.

No discussion of our reliability activity would be complete without mention of our programs presently underway to increase the professional capabilities available to Air Force Quality Control. Our program for the integration of quality control engineers into the Air Force Quality Control Representative's staff at plant level in the ballistic missile program is now well known. We have made rapid strides in extending this program to other areas of the AMC Quality Control program, particularly in new weapon systems. While our major effort has been the development of the quality control engineering program, we have also raised our field capability in statistics in order to meet successfully the demand for such a capability created by such new methods as the AGREE Sampling Procedures.

Technically, we have been very active in six major areas. This includes assistance in reviewing contractor proposals for reliability testing programs, as well as the normal, routine review of the entire gamut of commodity specifications.

A major part of the Air Force SQC effort is presently being expended and directed toward reliability. The well-known research work of Professor Benjamin Epstein in life testing is currently being converted by Professor Epstein into a Reliability and Life Testing Handbook, working with quality control statisticians from the three services.

Other research and development areas are under review and study. Implementing portions of the AGREE Task Group Three Sampling procedures in specifications and contracts, and disseminating directions to field organizations for surveillance, is presently requiring expenditure of one man-year. The growing complexity of such statistical methods, coupled with the increased technical complexity of functional, simulation, and reliability testing has increased the problems of Air Force Quality Control surveillance effort over such contractor conducted testing in most of our present weapon system programs. We have developed several techniques designed to assist in these problems for use in the Ballistic Missile Program, and expect to extend their use to other programs in the near future.

We are also serving on several groups within AMC and ARDC in initiating and monitoring reliability study contracts which are primarily statistical in nature. One of these is oriented towards development of meaningful reliability prediction models for use in the various stages of the design and development.

The final technical area I would like to mention is calibration. Assurance of quality and reliability must be based on a system of accurate and reputable measurement, requiring precision test and measurement equipment which are calibrated with standards of known accuracy.

Laboratories are now being established at 168 Air Force Bases to calibrate precision measurement equipment. The standards of measurement being placed in these laboratories are calibrated through a system, the measurement accuracy of which stems from the National Bureau of Standards.

Two committees concerned with calibration are now active. One of these, the "Air Force Committee on Calibration and Repair of Precision Measurement Equipment," is concerned with solving problems associated with implementation of the Air Force program. The other, the "NBS-Air Force Working Group on Precision Measurements and Standards," is concerned with requirements for development of new measurement standards.

A survey is now being conducted to determine the adequacy of contractor calibration systems. Although not yet completed, this survey has disclosed that generally contractor systems are adequate. There are problem areas, however, which we are taking action to correct as rapidly as possible.

The Need for Reliability Specifications

I remarked earlier that Air Force Quality Control is presently assisting in the development of a management type reliability specification. The course and tempo of events point very strongly toward the need for such a Reliability Specification. Although, theoretically speaking, reliability is simply another quality characteristic, in practice it is obviously very difficult to control. It is frequently extremely difficult to arrive at definitive requirements and conformance criteria that are meaningful. Production quantities may be such that sampling to provide adequate assurance is not feasible. Tests costs are extremely high. As an extreme example, how many ballistic missiles can we afford to fire solely to assure conformance? More important than any of the above, in the final analysis, is the necessity for "in-process controls" on reliability. The Air Force cannot afford the luxury of relying solely on end-item testing if it is to have assurance that our inventory will receive reliable weapons on an economical and timely basis. This is entirely consistent with present Air Force quality control philosophy that "in-process" type controls reduce the risk of delivery of unsatisfactory quality weapons with a resultant reduction in the capabilities of the operating commands. Enough knowledge currently exists to write a Reliability Program Specification. The Report of the Advisory Group on Reliability of Electronic Equipment (AGREE) and the Proposed Reliability Monitoring Program for Use in Design, Development and Production of Guided Missile Weapon Systems (ACGMR Report) are but

two outstanding examples of work done toward such a general reliability program specification.

Such a specification has already been drafted as part of the ARDC-AMC plan to implement this latter Report. While this specification is still in the coordination stage, it is reasonably safe to assume that it will require contractors to have a written, time-phased reliability program. It will probably include requirements for reliability requirement studies, design testing programs, design review and evaluation, and manufacturing test programs, as well as the traditional qualification testing, quality control system, data feedback system, and supplier quality control programs.

These requirements are stated in general terms to provide the necessary flexibility to adapt to differences in our various weapon system programs. A concise way for quality control personnel to visualize the general approach taken in this document is to consider its relation to reliability as analagous to the relation of MIL-Q-5923 to the Air Force Procurement Quality Control Program. The Plans and Programs Control Office, ARDC Detachment #1, located at Wright-Patterson Air Force Base, is the office of origin of this specification. The coordination version was drafted by that office with assistance from BMD; the Quality Control Office, Hq AMC; and the Assistant for Quality, Ballistic Missiles Center. An essential follow-on to such a document, of course, is the delineation of Air Force methods of surveillance of those portions of the contractor's reliability activity which naturally fall within our quality control mission.

There are, of course, additional reliability specifications, more technical in nature, which are necessary to complete the reliability requirements picture, and work in this area is also underway. However, time does not permit more detailed discussion of this activity here today.

Future

Much work, of course, remains to be done. I am sure that we are not so knowledgeable, in a field in which so much remains to be developed, that we can accurately assess all of the major areas of reliability in which we will be active, even in the immediate future. I feel it would not be remiss, however, to mention several areas that we feel are in need of greater attention as soon as possible. The first of these is simply the familiar need for indoctrination and education. While we have made a considerable effort here, much remains to be done. Currency alone is difficult in such a dynamic, changing field.

The second area is continued effort in maintaining and improving management teamwork. Here again, as I have previously stated, continuous management effort and attention in assuring a fully organized and coordinated approach to the many facets of reliability is essential to achievement of timely and economical solutions of the reliability problem.

A third area is exploration of development of new, and potentially radical, statistical models. Often, too much effort is expended toward

bending the real life situation to fit inappropriate, but available, statistical methods, in lieu of the more positive approach of developing new methodology appropriate and isomorphic to our true needs. As examples of my point, let me remind you that few statistical methods actually utilize all the facts that are in reality available concerning all pertinent variables. How can we integrate development testing results into production testing models to effect greater economies? What models are available which can reasonably be applied to development testing in which each successive item, and its test objectives, invariably are different? What models are available which fit the present concepts of reliability "growth"? Many other such questions can be posed, and we know that all too few solutions are available. Present time-cycles in the research to standardization of statistical methods sometimes lag behind the time-cycle of the "hardware" areas to which they could greatly contribute.

A fourth, and well-known area, is data feedback. The magnitude of this problem is universally recognized. I will mention several specific areas in which we feel aggressive; management action could potentially improve upon the existing situation. These are the earliest possible planning of the data feedback system, the earliest possible implementation of the system, and the allocation of sufficient analysis resources to assure immediate action on the failure information as soon as received by the analysis function.

Finally, we believe that the logistic aspects of maintenance and supply may deserve greater attention and consideration than have been given to date. Without delving deeply here, it would appear that an 80% reliable weapon needs roughly <u>twice</u> the field support that a 90% reliable weapon requires. This is obviously oversimplification, but serves as an indication that more facts in this area offer considerable potential resources savings.

It is, of course, possible to continue at great length in a similar vein, listing many more areas of consideration. Suffice it to say, in summary, that we in Air Force Quality Control, expect to continue to increase our emphasis on areas affecting reliability in the foreseeable future.

DEFENSE INSPECTION POLICIES AS APPLIED TO MILITARY SUBSISTENCE

Gordon J. Keefe
Supervisory Inspection Specialist
Office of The Quartermaster General

I. Background

a. Military subsistence procured by the Military Subsistence Supply Agency is inspected by three separate activities:

(1) the Veterinary Services (Army and Air Force),
(2) the U. S. Department of Agriculture, and
(3) the Military Subsistence Supply Agency

b. The Veterinary Services are responsible for sanitary inspection of all food establishments supplying the Department of Defense and quality inspection of animal-origin products in most areas of the United States. The Veterinary Services perform approximately 42% of the inspection workload for the Agency.

c. The U. S. Department of Agriculture performs inspection on all fresh fruits and vegetables, grain, most processed products and animal-origin products in specified areas in the United States. The U. S. Department of Agriculture performs approximately 46% of the Military Subsistence Supply Agency's inspection workload.

d. The Military Subsistence Supply Agency has civilian inspectors to perform inspection on operational rations, coffee and tea. The Agency performs approximately 12% of the inspection workload, directs and coordinates the over-all program.

e. Under the inspection philosophy that was in usage after World War II until the Spring 1957, the Government performed product inspection on lots tendered for acceptance using the attribute sampling tables established in Military Standard 105A. Defects classifications and acceptable quality levels were established either in the product specification or the contract. The contractor was obligated to maintain an inspection system acceptable to the Government. However, the burden was placed upon the Government to determine conformance of the products to the applicable specifications. This entailed visual and dimensional inspection in the contractor's plant and acceptance testing usually was performed at a Government laboratory. Most contractors at that time were unwilling to ship without receiving acceptable test results from the Government laboratory. This, at times, caused delays in meeting established delivery schedules and tied up essential warehouse space. This system of inspection raised questions of responsibility when defective material was discovered after acceptance. In the absence of fraud, if defects other than latent were encountered, the contractor had only a moral obligation to stand behind his product. Further, Government inspection tended to duplicate many of the examinations and tests performed by contractors. Optimum use was not made of contractor quality control systems.

II. Policies

The Department of Defense recognized the above deficiencies and issued a policy statement during 1957, requiring that "the supplier shall be required to perform examinations and tests set forth in the specifications to substantiate conformance of supplies to specification requirements". 1/ The Army Quartermaster Corps implemented this policy by directing the Military Subsistence Supply Agency as follows:

"Bid solicitations, and resultant contracts will require contractors to perform or arrange for the performance (as a service included in the contract price) of the examinations and tests set forth in the contract and/or specification quality assurance provisions on each item or lot of items offered to the Government for Acceptance". 2/ These policy statements have now been reduced to contract clauses which spell out in detail exactly what will be required of the contractor. In each case the contract will contain defects classification charts and associated acceptable quality levels, required points of examination (intermediate and end item), and complete instructions for testing. The contract will also prescribe the general format for inspection and test records. Under the terms of the above stated policy, the contractor may:

a. Have the U. S. Department of Agriculture perform all examinations and testing,

b. Perform the examinations, and sublet the testing to a commercial laboratory, or

c. Perform all examinations and testing in his own facility. Regardless of the option he selects, all examinations and tests must be performed and reported exactly as required by the contract. It is also very important that the contractor take action as the result of these examinations and tests. If rejection is required, this action, should be taken and corrective action initiated.

III. The Government Verification Program

The Government will perform verification examinations and testing to determine that contractors are performing the inspection as required by contract. The amount of verification inspection performed depends upon the reliability of the contractor inspection and the quality of the product. There are three types of verification inspection applicable to food.

1/ DOD Instruction 4155.8 dated 7 May 1957, Subject: Department of Defense Procurement Inspection Policies and Procedures for Items Covered by Military and Federal Specifications.

2/ QMC Circular No. 98 dated 8 September 1958, Subject: Contractor Responsibility for Inspection of Items Covered by Military and Federal Specifications. Superseding QMC Circular No. 17 dated 12 February 1958.

a. Type A - Verification inspection at the same rate as performed by the contractor - usually amounts to lot-by-lot inspection. It is used at the beginning of contracts when no quality history exists or when noncomparability between contractor and Government inspection results are established. It is obvious that it is the least desirable condition. If the Government is required to perform this rate of verification inspection indefinitely, the contractor is not complying with the terms of his contract. It is the intention of the Government to rely upon contractor generated inspection results to the maximum extent possible.

b. Type B - Verification Inspection - It amounts to the reduced form of inspection under MIL-STD-105A or permits the use of skip-lot techniques. When the skip-lot procedure is used, the number of lots skipped depends upon the establishment of comparability between the Government and contractor inspection results for a stipulated time interval. A contractor who establishes high reliability may have a minimum number of lots subjected to verification inspection. Process Quality Control Charts are a material aid to the contractor in maintaining high reliability at required quality levels. This type of verification inspection is expected to be attained in all contractors' plants.

c. Type C - Verification - The minimum form of verification inspection applicable to food inspection. Contractors eligible for this type of inspection are required to have and to maintain a quality control system. This system must be reduced to writing and verified by the Government. Verification by the Government usually amounts to inspection of product and records on a cyclical basis. In some cases the inspector may visit the plant only once during a two-week period. The contractor is relied upon to assure the Government that his product conforms in all respects to the terms of the contract, and is permitted to ship without the presence of an inspector. As long as the contractor keeps his system in control he need have no fear of nonacceptance by the Government. In the event that control is lost, it will be necessary for the Government to increase the severity of verification inspection by employing one of the other types. When control is reestablished, the contractor can again be considered for Type C verification inspection. The contractor who has and maintains an adequate quality control system satisfies the program objectives. Under such circumstances, Government inspection may be reduced to Type C with consequent economies. Statistics available from the National Canners Association indicate that approximately 67% of all food processers have quality control systems. It is recognized that some ofthese firms have varied quality control organizations, but nevertheless constitute resources which should be used to the maximum extent possible. The Military Subsistence Supply Agency will advise those food suppliers desiring to establish quality control systems for production of military subsistence.

IV. Safeguards

In addition to the verification inspection performed in the contractor's plant, the Military Subsistence Supply Agency places a warranty clause in each contract. The warranty clause permits the Military Subsistence Supply Agency to recover in the event that nonconforming supplies are received. It assures the unsuccessful bidder that all terms of the contract will be enforced.

V. Summation

The new policies implemented in the field of military subsistence will accomplish four objectives:

a. Eliminate dual responsibility of the Government and contractors with respect to examination and testing of subsistence by clearly making contractors responsible for this function and placing the Government in the role of verifying results.

b. Make maximum use of contractor quality control systems.

c. Improve the quality of supplies.

d. Minimize the cost of purchasing military subsistence.

The food industry has made considerable progress in adaptation of quality control techniques to its processes. Much work needs to be done toward developing quality control systems for the small producers. The field is unlimited in the application of quality control techniques to the production, storage and maintenance of foods. The success that the Government has with this program will depend to a large extent on the ingenuity of the quality control engineer in the food industry.

APPENDIX

CHART 1 - Mission of QMC

CHART 2 - QMC Inspection Program

CHART 3 - Inspection Organization, Military Subsistence Supply Agency

CHART 4 - Major Functions of the Inspection Division, Military Subsistance Supply Agency

CHART 5 - Comparison of Subsistence Inspection Programs

CHART 6 - Army-Industry Quality Control Program

CHART 7 - Experience Data - Contractor Testing Program

U. S. ARMY QUARTERMASTER MISSION

ARMY REGULATION 10-5 PARAGRAPH 48
22 May 1957

The Quartermaster Corps "develops, provides, and services food, clothing, petroleum, aerial supply equipment, other Quartermaster equipment and supplies, and provides Quartermaster services for the Army, and, as assigned, for the Navy, Air Force, and Marine Corps, and for foreign aid programs."

CHART 1

QUARTERMASTER CORPS INSPECTION PROGRAM

MISSION

Assure adequate quality in all supplies procured and stocked, including maintenance.

FUNCTIONS

OFFICE, QUARTERMASTER GENERAL

Promulgate inspection policies

Exercise staff supervision over field inspection activities

FIELD INSPECTION ACTIVITIES

Arrange or conduct:

Procurement inspection

Inspection of stocks in storage

Inspection of maintenance

CHART 2

Military Subsistence Supply Agency - Inspection Division

Mission: Operational inspection control of all central and regional procured subsistence for all military services.

Military Subsistence Inspection Offices: Ten offices located throughout the country under direction of Inspection Division, Military Subsistence Supply Agency.

Types of Inspection Performed:

Origin: Inspection performed at contractors' plants during process.

Destination: Inspection performed at destination for quantity and condition only.

Inspection Elements:

Veterinary Services: Consisting of officers and enlisted men performing origin and destination inspection of items of animal origin.

Quartermaster Corps: Utilizing civilian inspectors, responsible for origin and destination inspection of rations, coffee and tea, as well as component items peculiar to the rations, as assigned.

Department of Agriculture: Responsible for origin and destination inspection of fresh and processed fruits, vegetables, grains, and foods of animal origin as assigned.

CHART 3

INSPECTION DIVISION

MILITARY SUBSISTENCE SUPPLY AGENCY (MSSA)

MAJOR FUNCTIONS: (For all Single Manager owned and distributed Subsistence):

Arrange or conduct procurement inspection.

Develop quality control - quality assurance programs.

Prepare and maintain Inspection Handbook.

Prescribe inspection methods.

Manage laboratory testing programs.

Design and arrange for inspection instruments and accessories.

Prepare Section IV of specifications. (Quality Assurance Provisions)

Recommend specification improvements.

COMPARISON OF SUBSISTENCE INSPECTION PROGRAMS

	INDUSTRY	GOVERNMENT
ACCEPTANCE INSPECTION	Non-Systematic	Full inspection and testing of supplies to determine compliance with specifications.
	Systematic - Only a very small number of firms have an acceptable system.	Reduction-possible. Government withholds shipments pending acceptable results.
CONTRACTOR INSPECTION	Contract requires performance of inspection in addition to supply.	Conducts verification inspection or testing.
	Contractor may ship on his own results.	Reductions in inspection as soon as contractor achieves control.

CHART 5

ARMY - INDUSTRY QUALITY PROGRAM

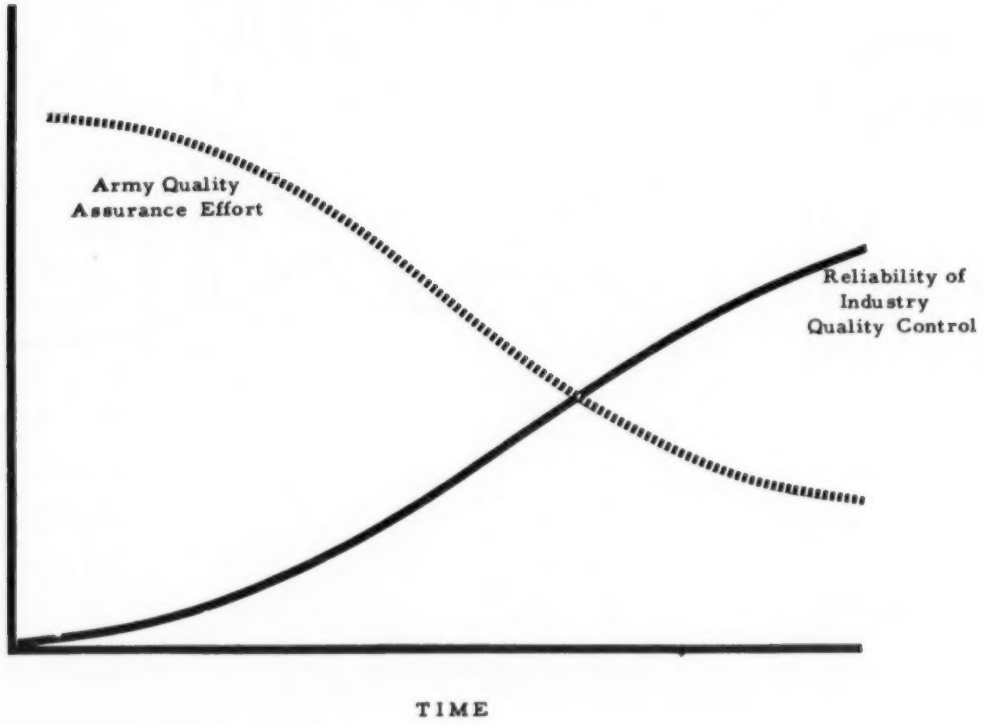


Chart 6

CONTRACTOR vs GOVERNMENT TESTING
QUARTERMASTER ITEMS
CONTRACTOR TESTING
GOVERNMENT TESTING
NUMBER OF CONTRACTS
T I M E

Chart 7

QUALITY CONTROL "FROM THE GROUND UP"

Harry D. Greiner
The Budd Company

Introduction

The transition from Inspection to Quality Control is taking place at our plant in an unusual manner. We call the change "Quality Control from the ground up" because it starts by having a large segment of the Inspection Department quality control conscious. This includes the hourly paid Inspectors. Because the story of the transition is unusual, it was suggested that it be recorded.

The First 2 1/2 Years

Eight years ago the Chief Inspector, who was approaching retirement, called to the attention of his inspection supervision that there was a new concept, called Quality Control, that appeared to have merit. He further stated that he could foresee the time when the Company would consider an outsider for a top position in Inspection if the responsible Inspection personnel had not become qualified in the field of Quality Control as well as in the field of Inspection. It was these comments that caused the writer to attend meetings of the American Society for Quality Control.

Trying to fully understand the various speakers at A. S. Q. C. meetings soon made it evident that some formal Education in applied statistics was necessary. The education process was started by taking the first course in "Masters Program in Applied Statistics". The text book used was "Statistical Quality Control" by E. L. Grant. Even though further courses in the Masters program were scheduled, it remained evident that "Statistical Quality Control" by E. L. Grant was the "Bible". It appeared that if a large segment of the Inspection Department understood "Grant" we would have a good foundation on which to build.

The Second 2 1/2 Years

About 5 1/2 years ago this idea of a broad base among the hourly Inspectors was discussed with the Personnel Department. It was decided at that time that any person who entered the Department would be told that"a working knowledge of Quality Control" would be necessary in order to ultimately attain the level of an "A" Inspector. This was reduced to writing and signed by each new member of the Inspection Department. "A working knowledge of Quality Control" was defined as a certificate covering a course in Q. C. from an accredited school (minimum grade 70) or the

ability to pass a written examination covering S. Q. C. by E. L. Grant. Final examinations used by schools within a 75 mile radius were secured and an Inspector who desired to take the test could pick one at random. He would be allowed three hours instead of the normal two hours. The use of notes and books is permitted. The examination is given on Company time.

It should be noted that the Quality Control requirement was not retroactive. This is important in the relations with the Union.

About this same time $\bar{X}$ & R charts without limits and histograms were introduced in the assembly areas.

The histogram data was coded so that it was possible to use it as a basis for process capability studies. Some histogram data was placed on modified nominal probability paper so that the percentage of non-conforming material and centering of a process was easily seen.

Concurrent with this chart activity, several noticeable changes took place in the shop. The first change was that equipment was adjusted on the basis of an average of 5 units instead of on a single measurement. The second change was that in the weekly Quality Meetings Production expected the Inspection Department to bring histograms and $\bar{X}$ & R charts. The third change was that Product Engineering and Methods Engineering began to receive and use process capability information for tolerance evaluation. At first the information was volunteered by Inspection; later it was often requested by Engineering.

During this period various members of Management, Product Engineering, Process Engineering and the Laboratory were invited to attend pertinent sessions of A. S. Q. C. Naturally all persons invited did not attend. However, the persons who did attend were interested in what they saw and heard. The reactions of these men, plus the many articles on Q. C. in various technical journals, further increased the interest of an already cooperative management.

While the seed of Q. C. was starting to grow in the minds of Production and Engineering, it came time to review and rewrite job classifications. The job classifications for Inspection and Laboratory personnel as now written contain two very important job requirements for an "A" grade.

(1) He must have a working knowledge of Q. C.
(2) He must have the ability to follow-up and eliminate complaints.

The first item was held to be applicable only to the men that entered the Department after this requirement was

made part of the new Inspector employee indoctrination.

The number of Inspectors taking courses in Q. C. was something above the area average during this period. In addition to this, we were able to obtain a man well advanced in an S. Q. C. Masters degree program to start with us as an hourly Inspector. His shop experience has been very valuable at the later stages of the program.

The 6th and 7th Years

About three years ago the point was reached where there was an opening for an "A" Inspector. The men with seniority and ability that were up for consideration had been hired after the start of the Q. C. requirement. Since this requirement was important, the Company up-graded, with Union cognizance, a man on the basis of an agreement that he take a course in Quality Control and pass same with a mark of 70 or better. He was to enroll the following semester.

This action served to highlight the question of Hourly Inspector Education. In the discussions that followed it became evident that the 2nd Shift Inspectors were not able to enroll in a non-credit course in Q. C. in the day time.

Sincere concern by management about the lack of non-credit educational facilities for 2nd Shift men eventually led to the formation of an unusual program. In this program the Company provided facilities for a 32 man classroom and arranged for outstanding instruction. To keep cost down two classes were scheduled. The men attended on their own time and prepaid their expenses. The 2nd Shift people attend for 1 1/2 hours prior to the start of the 2nd Shift and the 1st Shift people attend 1 1/2 hours after the close of the 1st Shift. The program was oversubscribed by Inspection and Laboratory hourly personnel. At the close of the 1st semester of the 2 semester program, management was so impressed that they refunded part or all of the tuition and book cost on "an accomplishment basis".

Concurrent with the educational program for hourly employees an Engineer was sent to Rutgers' 10 week course (one afternoon and evening a week). The purpose of this was to demonstrate to the Engineering Department the modern services in Quality Control that should be expected from the Inspection Department and/or Laboratory. This program is continuing.

Two years ago the Q. C. program for the hourly Inspection personnel had come close enough to our present goal to permit our taking a very constructive step. This step was that a Class "A" hourly man should be able by means of charts ($\bar{X}$ & R in particular) to analyze an operation in his area. The Inspector Class "A" job classification is interpreted to mean that the Inspector must be able to handle

production inspection problems.

Where an operation is in production (out of tryout) but the inspection is not on sampling, the Inspector must be able to identify the cause or causes in terms of man, machine, or material or a combination of these. He is expected to go over his findings (usually a chart) with the Production foreman. After he and the foreman agree on the findings, he is responsible to get the answers to three questions.

(1) What is the foreman going to do?

(2) When is he going to do it?

(3) What help is he going to need from Inspection, if any?

In the event of unsatisfactory progress the Inspector notifies his supervisor and the matter is handled between the Chief Inspector and the Superintendent. In the event of insufficient progress at this level the problem is handled between the Q. C. Manager and the Division Manager.

The 8th Year

At no time during the Quality Control program have we lost sight of the responsibility of the Production Department to produce quality. We have, however, in the last two years emphasized this responsibility in new ways. One of the methods used was to set our charts up so that we are recording operator reliability as well as machine capability. Production operators are required to check their own work and to note their findings on the charts posted at their operations. If the operator correctly checks and notes his work, the operation can be placed on a sampling basis within the specified plan.

Quality Control reports to management, on a weekly basis, the number of operations on sampling compared to the number it is possible to sample. Prior to issue of this report a sampling promise date, when applicable, is obtained from Production and is noted on the report. The charts as now designed have been a help to the Production foreman in his analysis of corrective action. Production has reduced the number of "process capability check requests"as the current system has been installed.

Conclusion

Building part of the responsibility for a level of quality that will permit sampling into the hourly inspector's job is very important. It must be done in a manner that makes it clear that the primary responsibility for

quality still lies with the man who produces the part. A program such as we have outlined requires patience and persistence. It has been worth all the effort and we now believe that we have one of the best and most fundamental programs in the metal trades.

THE RELIABILITY EFFORTS OF TEST ENGINEERING AT THE AIR FORCE MISSILE TEST CENTER

Dr. Albin N. Benson
Air Force Missile Test Center

The final quality of any product is frequently related to the effectiveness of the field testing. Therefore, it seems appropriate at this Annual Meeting of the A.S.Q.C. to describe briefly the Air Force Missile Test Center, which plays a most important part in establishing the performance and quality characteristics of our large missiles.

Our scientific knowledge is closely associated with the progress in experimental techniques. I need hardly to remind you of the important contributions made by Galileo, Archimedes and others in establishing the interrelationship between knowledge and experiment. In this day and age when the very life of our civilization depends on our scientific progress, it is essential that the quality of our defense weapons be established on the basis of utilization of adequate testing facilities. Slides will be used to give some indication of the effort being expended at the Air Force Missile Range in order to evaluate the various characteristics of missile systems. (Fig. 1). Amongst the most important of these characteristics is certainly quality and reliability.

There are 3 broad categories of personnel employed at the Test Center. First, the missile contractor or range user organization. These people are responsible for checking out, setting up, and firing their own missiles. In addition to contractors and sub-contractors such as Convair, Lockheed, Boeing, G. E. and Rocketdyne, there are also representatives of the governmental agencies directly involved in the design and manufacture of missiles. These include ABMA and Naval units.

The second broad category includes the range operating contractors who are under the supervision of the Air Force. As you probably know, Pan American Airways is the prime contractor and directly responsible for most of the ramifications of over-all management. They also handle the logistics support and are directly responsible for detailed, but important activities, such as the guard system at Cape Canaveral. The associate contractor for data collection and data reduction is RCA who is responsible for most of the technical operations. RCA personnel operate the radars, computers and photo processing installation. There are also a number of subcontractors within the PAA/RCA system.

The third category of operating organization at the Test Center is the Air Force made up of both military and civilian personnel. This group is headed up by Maj. General Donald N. Yates who has direct responsibility for all the far flung activities. Within the Air Force organization there is representation of SAC, Air Materiel Command and others. As missile programs progress into the operational stages, the using services enter into the evaluation picture and conduct test firings to establish the operational suitability

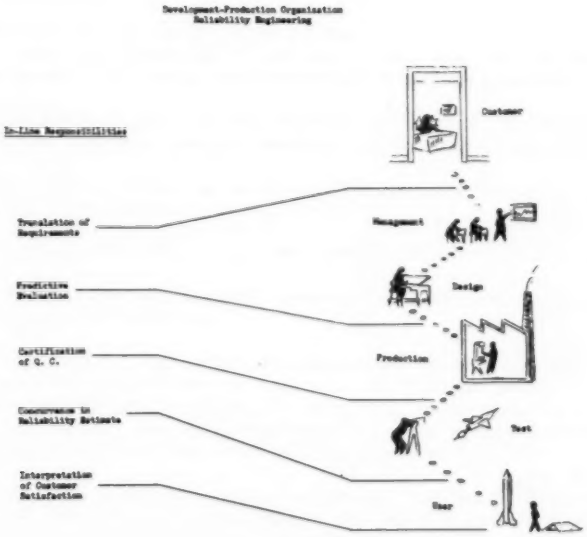
of the weapon systems under somewhat simulated service conditions. Slides will be shown indicating the relationship of the Directorate of Test Engineering to some of the other Air Force activities.

The Reliability Division is one of the units within the Directorate and I would like to explain briefly some of the things which we have been doing in an effort to promote the missile Reliability Program. Obviously, there are many facets to this work and for various reasons it will not be possible to touch upon all of them.

Some people may feel that the type of management organization is not important in relation to a Reliability or Quality Control Program. It has been stated "there is no organizational structure so bad that men of outstanding genius and loyalty cannot make it work with some degree of efficiency, however a sound organization aids men of genius to perform miracles and average men to achieve success." I think we are agreed that famous athletic coaches, such as Knute Rockne, have achieved success more through their ability to build a team than through their ability to produce outstanding individual players. With this in mind, we have, within the Reliability Division raised questions and encouraged discussion with missile contractors regarding the best type of organizational setup. Slides will be shown illustrating how the reliability program has been fitted into some typical organizations. In this connection, the question has often been raised as to what constitutes the work of a reliability engineer. At this meeting the question might well be generalized into a question as to the specific assignment of both the reliability and quality control engineers.

Slides will be presented to indicate a suggested breakdown into in-line responsibilities (Fig. 2), advisory responsibilities, and coordinating responsibilities. Considerable emphasis has been placed on the importance of design within the field of reliability. This is somewhat analogous to the frequently quoted quality control proverb that "quality cannot be inspected into a product but must be built into the product." There may, however, at times be a tendency to place too much of the reliability engineering effort within the Design Department thus in effect producing a design and a super design section and not giving consideration to the other aspects of reliability engineering. A series of slides are shown to indicate how the reliability program needs to encompass all phases of design, production, and use cycle.

U.S.A
AIR FORCE MISSILE TEST CENTER
3000 MILE RANGE
NORTH ATLANTIC
OCEAN
STATIONS
1 CANAVERAL AAFB
2 JUPITER AAFB
3 GRAND BAHAMA AAFB
4 ELEUTHERA AAFB
5 SAN SALVADOR AAFB
6 MAYAGUANA AAFB
7 GRAND TURK AAFB
8 DOMINICAN REP. AAFB
9 MAYAGUEZ AAFB
10 ST. LUCIA AAFB
11 FERNANDO NORONHA AAFB
12 ASCENSION AAFB
GULF OF MEXICO
CARIBBEAN SEA
SOUTH AMERICA
SOUTH PACIFIC OCEAN
EQUATOR
ASCENSION IS.
SOUTH ATLANTIC OCEAN


Development-Production Organization
Reliability Engineering
In-line Responsibilities
Customer
Translation of Requirements
Management
Predictive Evaluation
Design
Certification of Q. C.
Production
Concurrence in Reliability Returns
Test
Interpretation of Customer Satisfaction
User

FOOD PROCESSING DECISIONS BY MULTIFACTOR ANALYSIS

Amihud Kramer
University of Maryland

Any decision made regarding a single operation in the growing, manufacturing, or marketing of a food commodity, has its effect not only on the specific operation, but on every other operation in the entire cycle from the procurement of raw materials and supplies to the disposition of the finished product. This is merely another way of stating that any contemplated change in any given operation can be justified only after considering its effect on every other operation. This "total" approach is typical of Operation Research procedures. As applied to our problem, the item for which a decision is to be made is entered as a variable, while other items are entered as restrictions.

The working materials on which such decisions are to be based are accurate and precise yield and quality measurements, and accurate and realistic cost and price data. The following are examples of the nature of the information and equipment that are required:

1. Reliable methods for sampling and measuring quality of the raw material and the finished product, and their interrelationships.

2. Costs of procuring, or producing, the raw material; or gross returns expected by the grower of the raw product.

3. Relation of quality to yield of raw product in terms of tons per acre, or percentage of maximum yield, had the crop been harvested under conditions resulting in maximum yields.

4. Relation of quality of raw material to yield of the finished product, for each quality category of the finished product.

5. Total costs, direct and indirect for each manufacturing and marketing operation.

6. Estimated sales prices for each quality level of each commodity.

The following are some typical problems on which food processing management must make decisions:

1. Raw Product. - Should the raw product be procured or produced? Should purchase be made on a flat weight basis, or on a quality grade basis? If on quality, what premium should be paid for the higher quality? If several quality grades are available, which will be most profitable?

2. Processing. - Should defects be sorted out? Should they be sorted manually or mechanically? How much effort is justified to attempt to upgrade the product? What portion of the total

pack should go into what size container? What type container?

3. Sales and Pricing. - Can the price be reduced? If price is reduced can raw product payments be maintained?

A buyer's decision

Let us consider for our example, the problem of a canner of sweet peas, who is to contract with growers of peas for certain acreage, the crop of which he will be committed to accept, but for which he will pay on a quality grade basis. An accurate and precise means of measuring the quality of the raw peas is available. The decision which he faces is: how much can he afford to pay for the raw shelled peas of each quality level?* Thus his problem can be defined as one in which payment for the raw product is the variable, and all other conditions are restrictions.

Restriction (1) is shown in table 1 in terms of the prices which he can obtain for each quality level of the canned peas.

Restriction (2) is the total cost of processing and marketing the peas, less the cost of the raw material, as shown in column 3 of table 1.** The differences shown in column 4, are the sums available for purchasing the raw product.

Table 1. - Sums available for the purchase of raw peas.

1	2	3	4
Finished product grade	Anticipated price per doz. 303 cans	Costs** other than raw peas	Difference remaining for raw peas
Fancy, Grade A	$1.45	$1.01	$0.44
Extra, Grade B	1.30	1.00	.30
Standard, Grade C	1.15	0.98	.17

Restriction (3) is the relationship of yield of canned peas per ton of raw product, to quality of the raw peas. This relationship has been established, and is shown in table 2, columns 1 and 2, in terms of dozens of No. 303 cans per ton of peas varying in tenderometer grade.

---

* It is commonly assumed that since the tenderest peas are most likely to provide the highest quality product, that the highest price can be paid for the youngest peas that this is not necessarily the case, is demonstrated below.

** A profit item may or may not be included as part of this total cost figure.

Table 2. - Yield of canned peas per ton of raw peas harvested at different stages of maturity.

1	2	3	4	5	6
Raw pea tenderometer grade	Dozens of No. 303 cans per ton of raw peas				Indicated schedule of payment
	Total	Fancy	Extra	Standard	
80	166	166	—	—	73.04
90	186	186	—	—	81.84
100	195	185	10	—	84.40
110	205	62	143	—	70.18
120	212	21	191	—	66.54
130	217	—	195	22	62.24
140	222	—	155	67	57.89
150	227	—	91	136	50.42
160	230	—	46	184	45.08

Restriction (4) relates to the method of sorting the peas in the plant. Because of the availability of quality graders operating on the brine flotation principle, the field run peas are divided into different portion, in accordance with their density. Thus the younger and tenderer the peas, the more likely are they to float in a given brine concentration, so that a larger portion of the field run material may be packed as a high quality product. This relationship of the portion of the field run peas of different tenderometer grades falling into different quality categories has also been established, and is shown in table 2, columns 3, 4, and 5.

We now have the necessary information upon which to base a realistic decision regarding the value of the raw peas. Thus a ton of raw peas with a tenderometer grade of 80 is worth 166 times .44, or $70.04, since we can expect raw peas of this quality level to yield 166 dozen cans of fancy peas (table 2), and we have as much as $0.44 to pay for the raw material in one dozen cans of fancy peas (table 1). Similarly the price that can be offered for a ton of raw peas scoring 100 on the tenderometer, would be 185 times .44, plus 10 times .30, or $82.90. The complete schedule of payments which this processor is justified in offering his growers is shown in column 6 of table 2, and as the solid line in figure 1. If he should follow such a pay schedule and his estimate of prices and other costs is correct, he should break even, (or make a similar profit if a profit item is included in his cost figures) regardless of the quality of raw peas received.

This same approach can be made more general, and more flexible by introducing not only raw material quality, but other factors as variables rather than constants. Thus in tables 3 and 4, costs other than raw materials are also considered as a variable. In table 3, differences available for purchase of raw material are given for situations of varying production and marketing costs, and in table 4 are indicated prices which may be offered for raw peas of different tenderometer grade, when costs vary from $0.70 to $1.20 per dozen cans. Thus instead of

having a single curve indicating what the payments for the raw product should be, we now have a family of curves, as shown in figure 2.

Table 3. - Prices and costs, less raw material.

1	2	3	4	5	6	7	8
		Costs, less raw material					
Grade	Price	.70	.80	.90	1.00	1.10	1.20
		Difference, available for raw product					
Fancy	1.45	.75	.65	.55	.45	.35	.25
Extra	1.30	.60	.50	.40	.30	.20	.10
Standard	1.15	.45	.35	.25	.15	.05	—

Table 4. - Canner's break-even prices for sweet peas/ton, assuming following costs per doz. 303 cans, not including raw material.

1	2	3	4	5	6	7
	Costs/doz. 303, not including raw material					
Tenderometer	.70	.80	.90	1.00	1.10	1.20
80	124.50	107.90	91.30	74.70	58.10	41.50
90	139.50	120.90	102.30	83.70	65.10	46.50
100	144.79	125.29	105.79	86.29	66.79	47.29
110	132.23	111.73	91.23	70.73	50.23	29.73
120	130.38	109.18	87.98	66.78	45.58	24.38
130	126.95	105.25	84.18	63.11	42.04	20.97
140	123.21	101.01	78.81	56.61	34.41	12.21
150	115.77	93.07	70.37	47.67	24.97	2.27
160	110.40	87.40	64.40	41.40	18.40	—

A supplier's decision

Just as the canner is interested in avoiding operating at a loss regardless of the quality of the peas which are delivered, so is the supplier of these raw peas interested in obtaining a maximum return from his acreage. If he were to sell his peas on a flat weight basis, he would be inclined to harvest at the point of maximum yield, which is likely to be at a stage of maturity beyond that which will result in a 100 percent fancy pack. Furthermore, if he is to receive payment on a quality grade basis, he has further reason to investigate the quality yield relationship in order to reach a time of harvest which will yield him the highest return, since maximum yield may not necessarily coincide with a maximum return per acre.

Here again raw product quality is the variable, with acre yields, and production costs as the restrictions.

Restriction (1) is the potential yield per acre, had the peas been harvested at a stage of maturity coincident with maximum yield. In the example shown in table 5, this is given as 3000 lbs. per acre.

Table 5. - Grower's prices, per ton required to obtain the following gross returns per acre at different stages of maturity, assuming a potential yield of 3000 lbs/acre.

1	2	3	4	5	6	7	8	9
Tenderometer	% of Potential yield	$100	$110	$120	$130	$140	$150	$160
80	60	111.11	122.22	133.33	144.44	155.55	166.66	177.77
90	80	83.33	91.67	100.00	108.33	116.67	125.00	133.33
100	93	71.68	78.85	86.02	93.19	100.36	107.53	114.70
110	100	66.67	73.33	80.00	86.67	93.33	100.00	106.66
120	95	70.18	77.19	84.21	91.22	98.24	105.26	112.28
130	90	74.07	81.48	88.89	96.30	103.71	111.11	118.52
140	85	78.43	86.27	94.11	101.95	109.80	117.64	125.49
150	79	84.39	92.83	101.27	109.71	118.15	126.58	135.02
160	76	87.72	96.49	105.26	114.03	122.80	131.58	140.35

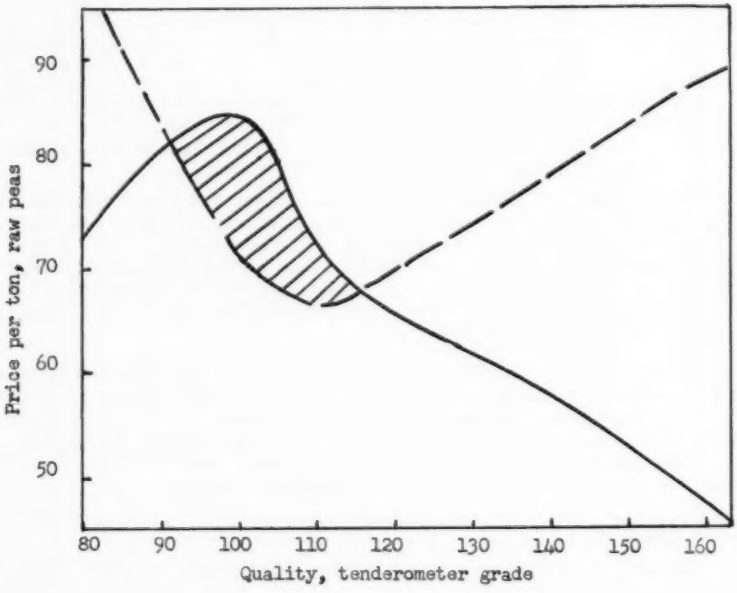
Restriction (2) is the relationship between the quality of the raw peas, and the percent of the maximum potential yield. This has been established, and is shown in columns 1 and 2 of table 5. Thus for example peas harvested at a very early stage of maturity so that they grade 80 on the tenderometer scale, yield only 60% of the potential yield, had harvesting been delayed for several days until a tenderometer grade of 110 had been reached.

Restriction (3) is the gross return per acre expected by the grower. If the grower is satisfied with a gross return of $100 per acre, he can afford to receive the prices per ton of peas shown in column 3, table 5, and as the broken line in figure 1. In the other columns in table 5 are shown similar schedules for increasing gross returns per acre, and these too are shown in figure 2 as a family of curves.

A best solution

Assuming that production of the raw peas as well as their processing into canned peas, and marketing, were all in the hands of one management, the question would not necessarily be: how much less to pay for more mature peas (canner's decision) or when to harvest for maximum return per acre (supplier's decision), but rather what is the best time to harvest peas so that they will provide maximum profits as a canned product. This then is a problem of maximizing profits, and minimizing costs, and has already been solved above, the answer being readily available from the more restricted problem shown in figure 1, the profit area being that where the two curves overlap. A profitable solution is therefore limited to tenderometer grade range of 90 to 115, and the best solution is at 102. Figure 2 provides similar information for a variety of canners cost and growers return conditions. The same data are presented in figure 3 in a form more in conformance with linear programming problems, with the best solution again being indicated at approximately 100 tenderometer. The data as presented in figure 3 can also be used to allocate costs as between farm and plant operations, when both are under the same management, and optimum harvests schedules cannot be maintained.

Figure 1. - Prices that can be paid by canner (solid line) or received by grower (broken line) per ton of raw peas of different quality.



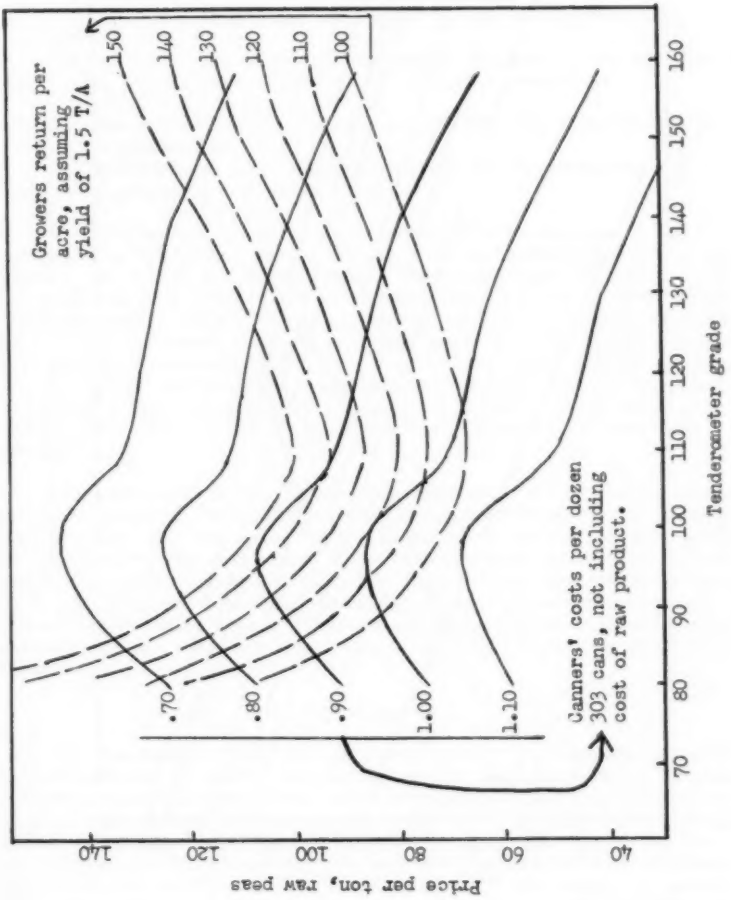


Figure 2. - Nomograph for indicating payment for raw peas on a quality grade basis, as influenced by growing, manufacturing and marketing costs.

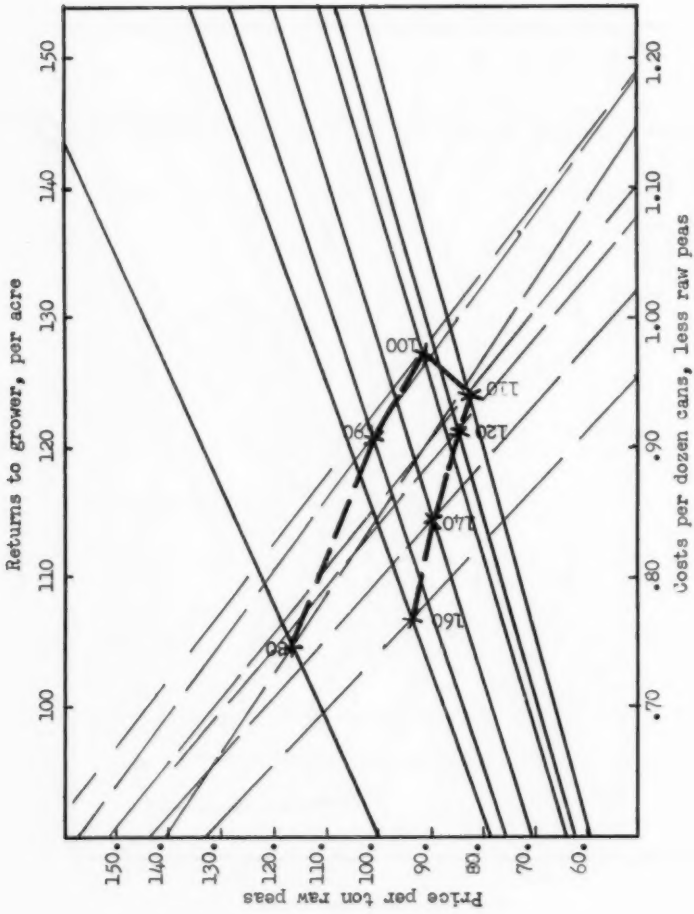


Figure 3. - Price per ton of raw peas of different tenderometer grade, considering canner's costs (solid lines) and grower's return (broken lines).

CAFETERIA SCHEDULING

Blair E. Olmstead
Senior Research Analyst
The Prudential Insurance Company of America

Two questions are almost always asked whenever a case history of statistical or any other "specialist" tools is presented:

1. Where does the specialist group fit in the company's organization?
2. How does the specialist group receive opportunities to practice its skills?

The Prudential, in addition to sales offices throughout the country, has 7 Regional Home Offices, 6 in the United States and one in Canada. An 8th is now being formed. Although there is some variation in size, the work done in these Regional Home Offices is essentially identical except that work of a corporate nature is concentrated in our Home Office in Newark. The Operations Research Unit, which conducted this project, is unique - that is, there is no other similar organization in any of our Regional Home Offices. Because its services are available to any group in the Company, it is part of the Corporate Office. The Unit is under the direct supervision of an Associate Actuary.

The work which we do is reported regularly, along with activities of other groups in the Planning and Development Department, throughout the Company. The quarterly report of Planning and Development activities contains simple two and three sentence descriptions of current projects; it is neither a collection of elaborate reports nor a carbon file. An item about another queueing project aroused the interest of one of our Regional Home Offices which was having trouble with lengthy cafeteria lines. We were contacted with the hope that we could assist in some way.

THE PROBLEM

Our Los Angeles Home Office, the first of the Regional Home Offices, was constructed in 1947. Our expansion has been so great since then (the Operations Research Unit was not around at the time to predict our growth) that it has been necessary to add a large wing to the building to house the additional people and equipment needed to service our policyholders. Unfortunately, it was not practical to add cafeteria facilities at the same time. The expansion in the number of people using the cafeteria, without expansion of the cafeteria facilities, brought on the queueing problem and, in turn, the Operations Research Unit.

The problem itself was deceptively simple. The cafeteria lines were too long. Employees were beginning to complain, with good reason, that they were spending too much time waiting in line to get their lunch. (It should be noted at this point, incidentally, that the Company administers its own commissary facilities, and virtually all employees eat in the Company cafeteria.)

Trained Methods personnel were available in Los Angeles to assist us in obtaining any necessary information. Since the benefits from an improved cafeteria schedule would be in the form of better morale rather than visible hard dollars, it was felt that the cost of sending a man to California should be avoided. If we had followed the operations research approach of having many experts in a variety of disciplines examine the problem, a substantial number of the analysts in our Unit would have been sent to California, bringing all work on their other projects to a halt. To avoid this, we cast about for some satisfactory alternative. Examination of the architects' plans for our other cafeterias showed that one of our Newark cafeterias had a physical layout that was quite similar to the one in Western Home Office. Unfortunately, it had three lines rather than the two in Los Angeles. Nonetheless, it looked like a good place to learn something about the general behavior of cafeteria lines.

To reduce the possibility of working toward a solution that would violate unknown statutes, we obtained legal advice about the restrictions which exist on serving meals in California. Three requirements which might affect our study were found:

> Women and minors may not work over five hours without a meal break of at least 30 minutes.
>
> A 10-minute rest period must be provided every four hours.
>
> A sneeze bar (partition up to a point above the normal level of sneeze) must separate the hot meals from the guests.

Since work starts at 8:15 A.M. in our Los Angeles office, the first restriction meant that women and minors must get through the cafeteria by 1:15. The 10-minute rest period every four hours created no problem. The sneeze bar was found to hurt service by muffling special requests and bringing the line to a halt while guest and server went "What?"... "What?" at each other.

As a Company practice, we would not want to feed anyone earlier than 11:15 A.M. The lunch period starting at 11:15 ends at noon and the 5 hour restriction means that the supper break for early lunchers would have to start at 5 o'clock when overtime is worked. Any earlier time would be quite awkward.

The cafeteria is located on the second floor of a 10-story building and occupies a substantial portion of that floor. It is served by six elevators, each with a capacity of 25. Normally, a guest arrives via elevator, walks down the corridor, and joins one of the two serving lines. After selecting his meal, he carries it on a tray to a dining area, where the tables are virtually all of a size to handle four guests. No attempt is made, however, to insist that all of the places at each table be occupied.

The usual menu consists of soup or juice, salad plate, side salad, hot meal, one or two kinds of sandwich, bread and rolls, a variety of beverages, and choice of dessert. Obviously, no one takes everything on the menu. However, selections range from a cup of tea or coffee,

just to keep one's friends company while they eat and you stay on your diet, to the three course dinner. Essentially, there are 14 service points in each cafeteria exclusive of the tray pick-up. The type of service varies from the self-service napkins and silverware, to drawn-to-order hot beverages.

THE INVESTIGATION

Despite all the work that has been done on queueing problems, there are no solutions for the multi-stage queues except for two station and a few degenerate three station cases. (A single stage queueing problem is one where complete service is given at one point and no further stops are required. When you select a line in front of a bank window to make a deposit, you are in a single stage, although possibly multi-gate situation, since more than one teller may be receiving deposits. However, when you make out your deposit slip at the high desks, another stage has been added. Queues may be found at both the desks and tellers' cages; hence this is a two stage, multiple gate queueing situation.)

How, then, were we to treat our 15 stage, parallel gate cafeteria line? Each service point could have its own queue, thus disrupting service at other stages. First of all we needed to learn about the "care and feeding" of cafeteria lines. This meant that data would be required; however, we did not know what data would be most useful. Accordingly, we decided to collect information experimentally in Newark, and then use the services of the Los Angeles methods men to obtain appropriate data. Whenever data is to be collected, it is desirable to know what questions it will be expected to answer. Too often valuable time and energy are wasted collecting large quantities of spectacularly useless information.

The next step was to determine what questions to ask. This can be a deceptively simple step - particularly if you know little about cafeteria lines except having participated in them.

The potential data which could be obtained about cafeteria lines is rather awesome. There seems almost no end to the number of ways this simple operation can be observed. Fortunately, however, a few statistics were obviously needed:

1. Some measure of the release rate of guests from each line.
2. Some measure of the input to each line.
3. Some measure of the holding time. (Holding time is the time required to service a customer at a stage - if we think of the whole counter as a single stage, it is the time the guests use to fill their trays, from picking up an empty tray to leaving the end of the counter to go to a table.)
4. Some measure of the distribution of the size of each line.

By treating the line as a single stage queueing problem, we were able to use somebody else's mathematics. By so doing we were successful in exposing what happened to the line before the first service

point. Later we examined what happened between the tray pick-up and release from the line.

We discussed cafeterias and cafeteria lines with those responsible for their operation in Newark. We were able to learn why many things were done, but unfortunately, they knew of no work that had been done in the field of cafeteria scheduling. (Since then, we have learned that several other companies have been studying this problem.)

Some data were already available. Particularly illuminating was the daily guest count by cafeteria serving lines. As mentioned above, the Newark cafeteria with which we were experimenting had three virtually parallel serving lines. However, two of these lines invariably served more people than the third. A little investigation revealed the cause: a few tables for six were available at the ends of the two more frequently used cafeteria lines. The most frequently used line was slightly closer to the elevators than the others. However, the second most frequently used line was furthest from the elevator arrival point.

After we had taken a few times and counts, we tried to analyze our findings. Our results were interesting but, initially, somewhat inconclusive. The lines did not behave the way our graphs and tables would lead us to believe. Luckily, the reason for this was extremely obvious: the people did not arrive in random order but rather at two-minute intervals as they were released for lunch. In fact, the intervals were not exactly two minutes because people came in elevator-size batches. Sometimes two elevators arrived simultaneously - other times there were gaps of almost a minute.

While we were testing methods of obtaining counts and times and recording data, we took action on another front. We obtained from Los Angeles the pattern of dismissal for lunch, and other information about the operations of their cafeteria.

Our test data had indicated that, under reasonably good conditions, the existing cafeteria facilities should be able to serve approximately 100 more people daily without much delay in line for anyone. Examination of the existing lunch schedule gave us the answer: as many as 130 people were regularly released at one time. Under the most favorable conditions, the last person of such a group must wait at least six and a half minutes before picking up his tray and entering the service line. Thus, approximately 10 minutes of his lunch period would be gone before he took his first bite. Unless care is taken, the following groups may suffer the same delay.

To find out more about batch-by-batch arrival behavior, a model cafeteria was constructed - on paper, of course, Although it could have been programmed for a computer, the manual methods proved to be quite satisfactory. This model cafeteria had three stations on each line. The first, corresponding to the empty tray pick-up point, was permitted queues of any length. Its input was generated from random number tables.

The second and third stations corresponded to the center (roughly hot meal pick-up) and end (full tray pick-up and leave) points of the serving counter. Each of these two stations were permitted queues up

to a set maximum. The inputs were received from the preceding points into their limited queues. If there was no room in the queue, the input was correspondingly reduced.

The average release rates from each station were identical, and set at a point slightly less per minute (the time unit used) than the permissible queue length for the latter two stations. The number released in each time unit were determined from random number tables, subject of course to the limitation of the number in the queue.

The model was validated by using actual input data. The holding times and first station queues were quite similar to those actually observed. We then used the model to test the effects of a variety of input schedules.

THE FINDINGS

It was found that the present serving facilities are such that, if people were released in groups of 14 each minute, less than 0.5 per cent would be required to wait more than one and one-half minutes before entering the line. However, if twice that number of people were released every two minutes, approximately 5 per cent would have to wait for one and one-half minutes and 1 per cent two minutes or more. This, of course, is a long way from the 1 per cent that must wait six and a half minutes when 130 people are released to two vacant lines.

While approximations of what would happen could be worked out statistically without the simulation of the model, the mathematics would be more costly and possibly inconclusive. A very real value of the model (which did not enter into this particular case history) is its sales appeal. Mathematical formulae filled with Greek letters and strange looking symbols do not have the model's ability to convince others that a thesis is correct. People can see the model work, whereas integral signs may leave them cold, or even scare them off.

We were now in a position to tell, in advance, just what would happen to the line in front of the tray pick-up point no matter what lunch release schedule was followed. This, however, was not the entire answer to the problem. A forecast of the number of employees five years hence had recently been made. Improvement in scheduling the arrival of guests would not be sufficient to provide for this expansion. It would soon be necessary to find some way to reduce the holding time - the time spent by the guests as they filled their trays. This meant that we would have to treat the cafeteria counter as a multi-stage queueing problem of a type that mathematicians have not yet solved.

Some of the 15 stages, it was evident, had a much smaller effect on the over-all holding time than others. Which should we measure--which could we eliminate? Merely studying the composition of the line could not tell us enough; we would have to know the history of each individual on line - when he arrived, whether he had made a special order, etc. Our ultimate objective was to produce a satisfactory recording scheme to be used by people we had never seen (the Los Angeles methods men) to supply us with data about an operation we had never seen. Therefore, we took times and counts to determine what the distributions were like at various stages of the line (e.g., release time

past the side salads).

THE RECOMMENDATIONS AND RESULTS

Examination of behavior at each of the various stages of the line gave us a fair amount of information about the proper order in which the service stops should be placed. Because of the equipment involved, the cost of rearranging a cafeteria service counter is considerable; but, fortunately, only a few changes appeared to have any value. Of these, one appeared to be extremely desirable; hot beverages, which were the last in line, were far and away the longest stop. Hot beverages are, of course, not ordered by everyone, but the probability of an order is not independent; people of similar sex and age tend to go through the line together, and the probability of a hot beverage depends quite substantially on age and sex. Other factors also enter the picture. For example, only a few orders for hot chocolate were received, all from one group of people; evidently no one else was aware that hot chocolate is offered.

In addition to recommending that the release times be set up on a minute-by-minute basis, a number of other suggestions were made. As mentioned above, the hot drink station was a major bottleneck. The capacity of the cafeteria lines would be substantially enlarged if an arrangement were developed that would permit some, if not all, of the people wishing hot drinks to move to one side and allow others to leave the line and go into the dining area. Our studies indicated that an "L" turn large enough to handle four would be sufficient to cause a marked increase in the capacity of the line. This recommendation has not been adopted because the expense is not presently justified. Nevertheless, it is one that will permit extension of the facilities at a cost substantially below that of installing a new cafeteria line.

Our second recommendation was to move such self-service items as silverware and napkins to points where a waiting line is likely to exist - in front of the hot beverages. The sneeze bar was found to have an adverse effect on communication with the hot meal servers, and a squawk box was suggested.

Not all of the recommendations were derived from the waiting line data. When we examined the stationing of the counter personnel, we found a tabulator whose function was to record the meal selections of the guests. This information was necessary for the kitchen to know what additional food would be required. It was found that another way of handling this was to develop "standard trays" of food (a standard tray would be 50 boiled potatoes or 100 slices of cake) and keep track of them. Knowledge of the rate at which standard trays were replaced early in the lunch period provide the kitchen with the data needed to insure that the late eaters were given the same meal as the early ones. This is a low grade illustration of another tool of operations research-servo theory. The "feed back" (a technical term) of food consumption permits the kitchen to have proper quantities of food prepared.

Solutions of problems that beget other problems are Pyrrhic victories. If, as a result of our efforts, people reach the dining area with full trays in their hands and have no place to sit, we shall receive no thanks. Accordingly, a survey of table usage was conducted.

It was soon found that the number of times a given table was used could not be increased appreciably. However, it was seen that a substantial number of the tables were occupied by only one or two guests. Thus, the purchase of tables for two would add the same effective number of seats as larger tables.

In addition to any inspirational or "how-to-do-it" values this case history may have, I believe it illustrates some points which are often ignored or overlooked. First, the problem and the investigating unit must be brought together. In this instance, if previous work on queueing problems had not been reported in an easily read form, the project would not have been handled by the Operations Research Unit. Second, it is not necessary to use the latest in electronic data processing equipment to make successful use of a model of the operation under study. A simple paper-and-pencil model is often sufficient for the required analysis and can be set up faster than a computer program. Third, the techniques associated with Operations Research can be employed profitably for small problems as well as large. Care must be taken, however, to ensure that O.R. and S.Q.C. do not become synonyms for applied statistics, with the result that the problems that such groups are permitted to pursue are restricted to mathematical aspects. Hidden restrictions may make certain mathematical "solutions" infeasible. In this case we had both legal and organizational restrictions. One must be aware of all the consequences of proposed solutions; be careful that the "solution" selected doesn't create new problems elsewhere. Restriction of the project to analyzing data will often cause one to overlook potential gains.

People with different backgrounds worked on this case: two "specialists" from the Operations Research Unit, a methods man who was on location and those responsible for the operation of the cafeterias. While certain ideas can be attributed to each person, the success of the project was a group achievement.

Not all of our recommendations were adopted, some because the expenditures were not yet justified, others because of "conservative thinking" (to us, that is) on the part of the cafeterias' administrators. However, enough recommendations were put into effect to produce a marked reduction in the waiting line. The evaluation received from the people concerned is that the revised system more than adequately meets their needs.

PROCESS CONTROL OF THE CONTINUOUSLY EXTENDED PRODUCT

TEXTILE MANUFACTURE

Charles D. Ferris
Mohasco Industries, Incorporated

Introduction

While the term "continuously extended product" is apt to first bring to mind such products as extruded metal and plastic materials, paper, and film, there can be no doubt that textile materials also qualify for this description. Literally miles of filament strands are produced from chemical solutions in a single run by synthetic fiber manufacturers. Wool, cotton, and synthetic staple fibers first assume continuous form when they are rubbed together into round, continuous strands. Long, unbroken lengths of yarn are then produced from this untwisted material in spinning and twisting operations. Finally, the yarns are woven together to form long lengths of cloth to which are applied various sizings. Non-woven fabrics, fabricated by pressing and bonding short fibers together, are also becoming increasingly common. Yes, the problems of quality control people in the continuously extended product field are indeed our's also.

I will attempt today to acquaint those of you in other extended product areas with a quality control approach and techniques which are both unique to the continuous product and are likely to have some application to your problems. I would hope that my remarks would profit the beginner in textile quality control but would imagine that the veteran in my industry would be familiar with most of the material.

Let's start with the quality of staple fiber in non-continuous form.

Sampling Heterogeneous Stock

In the case of raw material, little difficulty is encountered in evaluating the quality of the new synthetic fibers. While such fiber characteristics as fineness, average length, strength, and elongation will vary appreciably throughout a batch, the incoming material is reasonably homogeneous, and the level and variation characteristics can be readily estimated from a small sample. Real sampling problems arise, however, in the case of the old-timers such as wool and cotton. Here the variation in fiber quality throughout a bale and from bale to bale can be quite great, and it is necessary to resort to systematic core-boring to obtain a representative sample. The number of scoured pounds a shipment of greasy wool at dock side will yield is estimated through this means, for example. After careful analysis of within and among bale variation, sampling plans have been devised which specify that if one wishes to estimate the lot yield within plus or minus y per cent one must core-bore n bales k times, scour the material, and compare the scoured weight with the original weight.

Once the fiber has been accepted, it is scoured if of animal or vegetable origin and blended with other fibers to obtain homogeneous

stock with the desired running properties, color, resilience, toughness, etc. The blended stock may then be dyed - in which case it must be thoroughly reblended to assure a uniform color - or it may be immediately converted to yarn, to be dyed in package or skein form when needed for cloth manufacture. In any event the quality problem of greatest concern to the production and quality control man is the thoroughness of the blend. Is the stock well mixed? The textile industry spends considerable time and money checking on this point. Analysis of fiber content, fiber length and fineness, oil content, and color fluctuation throughout the yarn length and from yarn package to yarn package will indicate the thoroughness of the mix and the need for revision of blending procedures.

As regards matching a dyed stock blend to a standard color, samples are selected systematically from each kettle's poundage, mixed, and converted to yarn. If the match is off, additional stock is dyed a color which will bring the shade of the entire batch within tolerance limits after blending.

Determining the Quality of Extended Material

I believe that all parties concerned with controlling the quality of the continuous strand or strip will agree that one of the more frustrating aspects of the business is the difficulty of getting at the product. The traditional way of estimating levels and measuring variation is through random sampling - quite simple in the case of small, discrete products but not so easily accomplished with extended material. The approach in the textile field, which I am sure jibes closely with the methods employed in rubber, paper, and film, is as follows:

First determine the unit length or unit area of interest. Are we concerned with weight fluctuation throughout yarn within a foot, a yard, or 50 yards.

Second, measure the quality characteristic in a sufficient number of units throughout an extended length to determine the pattern of variation.

Third, determine whether this pattern holds in other extended lengths of the same type, e.g. other lengths of roving (continuous, untwisted strands), skeins of yarn, lengths of cloth.

Fourth, determine the manner in which the quality level varies from extended length to extended length.

Fifth, after studying such data determine the optimum sampling arrangement.

Such a sampling plan will contain one or both of the following provisions:

Cut from the end of a portion of the extended lengths produced and measure the characteristic of interest throughout the piece

removed in one test (weight of piece, for example), at one point in the piece, or at several points.

Cut a sample at several points throughout a portion of the extended length, i.e. at the start, a quarter of the way through, a half of the way through, etc.

Finally, determine the variability present in the material by studying the test data in conjunction with known variation patterns (the test data may always be systematically lower or higher than the material average due to sampling restrictions) and decide the disposition and corrective action warranted.

A good example of a continuous material in textile manufacture which varies in every conceivable way are the ends of roving produced by a card. The web delivered by the finisher card is divided into anywhere from 60 to 200 strips by tapes which "condense" the sliver into continuous, round, untwisted strands which are thereupon wound onto a number of spools. The density of the strands will vary appreciably longitudinally, from end to end across the spool, from spool to spool, and from card to card. The weight control sampling plan in this instance requires that a certain number of feet of the roving be pulled from a percentage of the finished spools after they are full, that the total weight of the ends removed from each spool be obtained, and that the interior of the spool be sampled if the outside weight is beyond the weight tolerance limits. The plan also specifies that the extreme and center ends on a spool be checked individually at intervals to determine the density variation across the card.

Other examples of systematic sampling in the textile area are carpet pile yarn and backsize weight checks which must take into account longitudinal and lateral variation throughout the cloth, from roll to roll, and from loom to loom.

Control Charting Techniques

Control charts are put to good use in textile quality control work, particularly in yarn manufacture. Yarn unit length variation is so great relative to shifts of interest that sample averages must be compared to control limits to determine whether the point is significantly higher or lower than the standard level. Roving and yarn weight charts are especially helpful in indicating the need for corrective action.

A word of caution with respect to the construction of a control chart. Make certain that sample values are independent. Do not make the mistake of computing the control limits on the basis of ranges of samples selected purposively from material characterized by systematic variation. For example, if squares are cut from the sides and middle of a strip of cloth, the weight of each square computed and control limits calculated on the basis of the average range of the three weights, the limits will be in error should the weight tend to be consistently high at one lateral point. The range will be artifically inflated due to the presence of the systematic variation, and the limits will not be

as sensitive to assignable causes of variation as they should be. The limits should be computed instead on the basis of an average range which has been reduced by the systematic difference in levels which is known to exist.

Yarn break frequency data should be collected in yarn finishing and at weaving. Such figures can shed much light on the quality of the blends, on the quality of the yarn construction operation, and can be plotted on a defects-per-unit-of-production type control chart.

Returning to variable data, if a small interval fluctuation is important, the textile quality control man can make good use of instruments which record the quality characteristic continuously as the material passes through the measuring head. Evenness testers which measure the density of roving and yarn on the basis of capacitance changes or bulk variation can be quite helpful in evaluating equipment capabilities - particularly when exacting flat goods appearance standards are involved. Other instruments - some employing radioactive material - have been developed to check cloth density.

As regards over-all measures of yarn variation, the coefficient of variation (the standard deviation divided by the mean) has been found to be a very meaningful way of keeping track of variability from one period to the next, of comparing machine performance, and of relating yarn variability to cloth appearance quality.

Role Re Process Specifications

You know, every so often I will be listening to a textile quality control paper or reading an article on the subject, and I will say to myself, "Why this is not quality control but process engineering. This speaker is talking about process methods and specifications." And yet as I ponder the message I realize that the remarks do definitely relate to a province of quality control for the speaker or the writer is discussing specifications which are developed by joining the production man's knowledge of the process to quality control's ability to design tests and throw relationships into sharp relief through analysis of data. This area of common responsibility exists in the textile field. The creative effort naturally resides almost entirely in the hands of the technical people, the production and methods men - and yet, although quality control does not say "install the machine" or "do things this way", he can be of assistance you might say in the second degree.

To be specific, what are the right blend specifications. To what extent can we mix fibers of varying fineness, maturity, staple length, elasticity, etc. What is the relationship of the various combinations with yield, yarn strength, frequency of neps, yarn appearance, production, and ability of the yarn to stand up in service. The relationships are all important. Literally hundreds of thousands of dollars ride on the blend make-up decision in some textile operations. The relationships can only be ascertained economically by designing efficient experiments and dredging up relationships from the welter of data.

In the machine setting area the quality control man can provide similar service. Take such a statement as this: "Increasing the licker-in screen length by 5/8" along the curved dimension reduces fly waste about .6% of lap weight." [1]

In the process specification area the design of experiments can be again put to work in determining, say, the relationship between type lubricant and warp defects.[2] Finally, let us not forget before leaving this phase of quality control that process capability studies can be of tremendous assistance in establishing reasonable mill limits on such characteristics as skein weight variation, yarn strength, and grease content.

Product Quality Ratings

Process control effort is often accelerated and facilitated by the distribution of quality rating information. A summarization of the finished product quality - in toto and in terms of the primary elements - can reveal unfavorable quality trends long before they would ordinarily come to the attention of mill personnel and serves to place component quality in its proper perspective. I should like to briefly outline two rating schemes which have been used to good effect in the textile field.

The quality of finished yarn is a function of a large number of characteristics which can be grouped into roughly five classifications: Physical characteristics, chemical characteristics, color match, running and visual quality. One rating arrangement recently devised first weights the relative importance of each category and then assigns ratings to deviations from specification and/or ideal performance within each category. Batch yarn weights, for example, are rated in terms of the deviation of the test average from the nominal density. The visual quality is dependent upon the appearance of the yarn in the finished cloth, the amount of reworking required, and the final classification. And so on. Finally, the ratings are summed and averaged to obtain category and over-all ratings. Thus, the quality of finished yarn can be expressed in terms of a single number such as 95, 70, 60, etc.

Summarizing the appearance quality of cloth has always been somewhat of a problem. Fabric blemishes have three dimensions: frequency, intensity, and size; defect characteristics which militate against a simple summarization of appearance quality. One rating scheme which is quite prevalent in the flat goods field involves the assignment of demerit points for intensity and size, the rating of each defect regardless of clustering or frequency, the summing of the demerits, and the obtaining of average demerits per square yard inspected by department responsible, type defect, and for all defects. Another rating method requires that the fabric be split into imaginary lineal units of a fixed size, that any defect condition within a unit be rated slight, minor, major, or critical and assigned appropriate demerits, and that the demerits be summed and expressed in terms of the average number of demerits per unit length inspected. An index of the poorest quality present can be obtained by calculating the percentage of the unit

lengths inspected which contain one or more critical defects. Fabrics can be rated very quickly by this method.

Responsibility for Process Quality

It is generally recognized in industry today that the primary responsibility for meeting process specifications is that of production, the people who make the goods, and that it would be a mistake to weaken this responsibility in any way. And yet Quality Control by definition has a secondary responsibility. It is my opinion that this secondary responsibility in the case of extended product batches is not as strong as that which exists in the case of discrete products.

I have reference to the relative feasibility of toll-gate checks by Quality Control. When products such as metal parts are involved, it is often quite practical for Quality Control to guarantee average outgoing quality or lot tolerance protection. A substantial quantity of the parts can be quickly checked while the lot is in temporary storage awaiting the call for the next operation, and the material will be accepted or rejected on the basis of the number of defective pieces in the sample. In the extended product field, unfortunately, the conditions favorable to this type of quality assurance often do not exist. Too frequently the quality control man is confronted with one or more of the following situations: the defect is hidden; the material is not easily sampled and/or tested; the product must be passed on to the next operation without delay; it is not practical to clear the product of the defective material even though the condition is detected before the batch is released. In the textile field, practically all these circumstances obtain as regards yarn irregularities; none of them apply in the case of a critical characteristic like color match.

Thus, it is quite clear that in textile manufacture the quality of material sent on to the next operation is entirely dependent on the job done by production supervision and the operators. If the latter are not quality-minded, we are often just stuck with the goods, and product quality and process costs suffer accordingly. Quality Control under these conditions must concentrate on long term quality. They must make certain that the control system is right and be alert to any signs that substandard batch quality will persist. By fulfilling this basic responsibility and emphasizing this phase of their work, they can be assured that they are preventing the manufacture of defective material and are operating constructively in their organization.

---

1. "The Control of Carding Weights and Its Influence on Yarn Quality," J. F. Brogden, Textile Quality Control Papers, Vol. 2, published by the Textile Division, A.S.Q.C.

2. "A Designed Experiment to Evaluate Seven Yarn Lubricants," H. M. Hill and D. Wheeler, A.S.Q.C. National Convention Transactions, 1958.

WHAT A.S.Q.C. IS DOING FOR THE QUALITY CONTROL ENGINEER

Leon Bass
Manager, Quality Control
Jet Engine Department
General Electric Company

When the American Society For Quality Control was organized thirteen years ago, the aim of the newly-formed society was to promote the use of scientific techniques for control of quality in industry. These techniques were supported by the Armed Forces during World War II, and at the end of the war there was danger that they would be dropped due to a lack of general appreciation of their capabilities. Fortunately, among the users of these techniques there were men of vision who felt the need of a common meeting ground, so this group of men organized the American Society For Quality Control to fulfill that need.

The success of this effort is evidenced in a number of ways. The Society has grown from a group of about 1200 Charter Members to today's membership which exceeds 11,000 - including 105 Sections and 6 Divisions. These numbers indicate how completely the methods have been accepted - in principle if not always in practice - throughout industry. The technology itself has expanded into greater and greater territories from the original (and still useful) techniques of $\bar{X}$ and R Charts and Attributes Sampling Plans, to a variety of tools including adaptations of the old techniques to more special or more complex problems, new techniques to solve the original problems more efficiently (Sequential and Continuous Sampling Plans, Median and Quasi Ranges), and techniques for problems which were not originally a part of the Quality Control area of interest; for example: Vendor Rating, Machine and Gage Capability, Reliability, and Experimental Techniques.

While the methodology of Quality Control was being developed, perhaps an even more important change was taking place in the building of a philosophy for control of quality, changing from a scope encompassing manufacturing and measuring activities alone to one encompassing just about every activity in industry. The result of this has been promotion of the status of the Quality Control Group from that of a service organization to that of a dynamic force in the conception, development, marketing, production, and distribution of industrial products.

Quite naturally - in keeping with this increase in numbers, this development of a variety of tools to solve a wider and wider variety of problems, and this expansion in scope and level of activities - there has evolved a new, important group of men on the industrial scene. These men are the Quality Control Engineers - the professionals of the science of Quality Control and the experts on the application of statistical theory to shop and quality problems in general.

Since the existence of the Quality Control Engineer is, in some measure, due to the activity and efforts of the American Society For Quality Control, it is quite proper that A.S.Q.C. should be regarded as the professional society for the Quality Control Engineer. This thought has been pursued by some even to the extent of suggesting that the name be changed to the American Society of Quality Engineers, or something similar.

On the other hand, there is a feeling in some quarters that there is not enough "meat" in Quality Control Engineering to warrant a special title at all. This feeling is generally supported by the proposal that some one or another of the several existing professional societies includes the Quality Control Engineering field as a special area of interest. These societies, either officially or unofficially, believe that the Quality Control Engineer belongs in their fold. This argument is generally advanced by people who have limited knowledge of Quality Control and the scope of its application. It is true, of course, that some functions of other societies do overlap into specific and limited areas of Quality Control responsibility, but it is just as true that full consideration of quality problems takes us into areas of industrial, mechanical, chemical, or electrical engineering. When the quality problem leads us into the design of the product, the true Quality Control Engineer must understand enough of the principles of mechanical engineering, for example, to be able to evaluate alternate solutions, test them, and pattern control measures in keeping with the solution. When the problem involves measure of fluid flow or chemical yield, he must understand enough of fluid dynamics or chemical engineering to evaluate measurement equipment; when the problem involves vendors, he must understand procurement; and when it involves efficient operations, he must understand industrial engineering. To all these problems, he must apply variously the total quality control concept; the statistical methodology of acceptance and control; design and analysis of experiments; and the principles of measurement - including a detailed familiarity with measuring equipment and its capabilities. The American Society For Quality Control has played such an intimate part in the development of the technology and the profession, that I firmly believe that A.S.Q.C. alone can possibly provide the services for all phases of his work that the Quality Control Engineer should expect to receive from his professional society. The Quality Control Engineer, then, should regard A.S.Q.C. as his professional society.

If the American Society For Quality Control expects to serve as the professional society for this group of engineers, it must then take a critical look at its contributions to the nurture and development of the profession. The Society must first review its current activities and services; it must decide if each activity is needed, and remove the ones that are not; it must revise those services that need re-orientation and improvement; it must decide what additional services are required and undertake their development so that the professional society function is complete.

Realizing that I cannot do more than a cursory job, I would like to outline the functions which have been, and are being, provided to the Quality Control Engineer by A.S.Q.C., its Divisions and its Sections. After that, I would like to suggest some areas where additional effort needs to be applied. The provided functions are:

I. TECHNICAL INFORMATION

The first thing that comes to mind is the preparation and dissemination of technical information. Here A.S.Q.C. has made a most extensive contribution, for a great deal of effort has been spent in exchanging information, experiences, and methods. This, of course, is of primary importance since most colleges and universities have not yet included sufficient industrial statistics and quality control in their curricula.

We hope to see the day when such courses will be part of the undergraduate program for engineers, but until that day comes, the main load of distributing Quality Control technical information will continue to fall to A.S.Q.C. We have four distinct media through which technical information is passed on to the Quality Control Engineer:

1. Periodicals

You are, of course, familiar with INDUSTRIAL QUALITY CONTROL, the official organ of A.S.Q.C. This magazine originally started in 1944 as a publication of the Society of Quality Control Engineers in cooperation with the university of Buffalo. Published bi-monthly by Dr. Martin Brumbaugh, it was adopted by the American Society For Quality Control in March, 1946. Besides all the Society news and information, there are technical papers published each month. Effort has been put forth by the Editorial Board, under the leadership of Dr. Mason Wescott, to achieve a balance between the "practical" and "theoretical" in published articles, and at the same time to maintain a consistently high standard of technical quality. Some of the articles are case studies of successful applications involving simple methods, while others may be specialized sampling or control techniques, or sophisticated curve fitting or response surface methods. All of them recount experiences or display methods valuable to the Quality Control Engineer. As evidence of the fine job done, INDUSTRIAL QUALITY CONTROL became a monthly publication in January of 1955.

Use of the more erudite statistical methods has become increasingly widespread in the Quality Control field. This is particularly true in the chemical field where designed experiments are used in many applications. This has led to a demand for more articles of greater mathematical and statistical content, and to the formation and publication of a new journal called TECHNOMETRICS. Initially, at least, TECHNOMETRICS will be published quarterly; co-sponsored by the American Society For Quality Control and the American Statistical Association. Articles will relate to problems of Quality Control, Reliability, and Experimentation. This new project should get off to a good start with Dr. J. Stuart Hunter at the editor's desk.

2. The Milwaukee Office

The Milwaukee Office also handles publication and distribution of technical information. Sometimes a particular aspect of Quality Control creates a demand for reprints of INDUSTRIAL QUALITY CONTROL articles, and a supply will then be kept on hand for sale at a modest price. Some of the more important publications have been Dorian Shainin's "Lot Plot Plan", Leonard Seder's "Span Plan", and more recently, the Charles R. Hicks articles on "Analysis of Variance". Sections, Regions, and Divisions may use the services of the Milwaukee Office to distribute conference transactions, handbooks, and special instructional or informational material. Sometimes another technical or industrial organization will publish literature relating to Quality Control and the Milwaukee A.S.Q.C. Office may handle distribution and sale to the A.S.Q.C. Membership. Examples of this are the ASTM's "Manual on Presentation of Data" and the NSMPA's "Quality Control Series". Also of importance is the A.S.Q.C. "Standards Series" prepared under the direction of Harold F. Dodge. A stock list of A.S.Q.C. publications is periodically published in INDUSTRIAL QUALITY CONTROL, and I suggest that you review it to see what is available.

3. Section Meetings

A third means of transmitting technical information is through Section meetings. These may be the regular business meetings where the featured program of the evening is often a speaker on a technical subject, or they may be specific meetings for the purpose of presenting technical papers or conducting technical seminars. In many ways, the Section meeting is the most effective way of all to reach the Quality Control Engineer because it is accessible to practically all of the Section members, and the usually informal atmosphere is conducive to more discussion. In order to assist the Sections in the preparation of good programs, A.S.Q.C. publishes a Speakers List and offers, through Area Program Representatives or the District Representatives, detailed assistance in programming, as well as in obtaining speakers and other aids. This service is not used nearly so much as it should be; this is due partly to the fact that Section Program Chairmen are not aware of it, and partly to the fact that the service is relatively new.

4. Conferences and Symposia

In addition to these media, each year there are many conferences and symposia which afford an opportunity for the Quality Control Engineer to exchange ideas and methods with the others in his profession. This National A.S.Q.C. Convention which you are now attending is the oldest and largest of these, and offers the broadest coverage. The Divisions usually have conferences - generally national in scope - to present papers applicable to their industries and areas of interest. The Regions, and the Sections, on the other hand, may present a conference that is aimed at reaching a larger proportion of local or semi-local members who may find it difficult to attend the National Convention or the Division Conferences. The growth in the number of, and the attendance at these conferences indicates how well they serve the membership.

II. EDUCATION AND TRAINING

The second area where A.S.Q.C. serves the Quality Control Engineer is in education and training. Because of the "bootstrap" nature of Quality Control application, colleges and universities have not been adequately prepared to handle the training of Quality Control Engineers. In order to fill the gap, A.S.Q.C. has sponsored or coordinated numerous training courses.

A great many Sections have an educational program as a part of their Section activities. Of necessity, they are often elementary courses limited to discussion of the uses of control charts and sampling plans, (this is a factor that causes many outsiders to feel that the basic content of Quality Control methodology is weak) but even these provide a necessary first step toward a full understanding of the science. Likewise, many conferences provide a "quickie" course in Quality Control, and some offer education in the related topics of analysis of variance and reliability.

In addition to these classes, there are some courses sponsored by Sections and Divisions - often in conjunction with colleges and universities - that are of a more advanced nature. Some of the larger Sections are offering fairly erudite courses in statistics. The Chemical Division is quite successfully presenting courses in the Box and other techniques while the

Electronics Division is teaming up with the Institute of Radio Engineers to organize and conduct a course in Reliability which is to be presented for the first time this fall.

Without doubt, these educational programs have had - and will continue to have - tremendous effect in raising the general competence of our Quality Engineering personnel. It is recognized that this type of training program is not the most desirable for supplying industry with the Quality Control Engineers that it needs, and that eventually the bulk of the training should be done in the college classroom. However, I shall say more about this a little later.

III. STANDARDIZATION

A third area of assistance to the Quality Control Engineer is standardization of terminology and symbols. Historically, one of the most troublesome obstacles to gaining understanding of the subject has been notation. Thus far, the American Society For Quality Control has published two "Standards Series" - one for control charts and another for attributes sampling. Others are "in the mill". In addition, A.S.Q.C. has cooperated with the International Committee on Standardization to obtain agreement in notation outside the United States. This work is pushing ahead, and the eventual gain will be to the engineer who will be able to communicate his experiences more freely and easily, and who will be able to read his technical literature without reference to extensive glossaries accompanying each paper.

IV. PROFESSIONAL DEVELOPMENT

A fourth area of service to the Quality Control Engineer - and one of which you may not be fully aware - is Professional Development. By this I mean those A.S.Q.C. activities which are deliberately aimed at improving the standing of the Quality Control Engineer in the eyes of his employer the technical people he works with, and the public in general. For the past year-and-a-half, an ad hoc committee under the direction of Dr. A.V. Feigenbaum has been investigating this problem from every possible standpoint, and the committee has recommended that a permanent organization of five committees reporting to the Vice President for Professional Development, be formed. The five committees will work in these areas:

1. Committee on Development of Work Elements

The purpose is to define the nature and scope of the duties of the Quality Control Engineer. A list of work elements has been compiled by the special committee, and was published in the February, 1958 issue of INDUSTRIAL QUALITY CONTROL. This list will be studied and revised until general agreement is reached that it is an adequate description of the duties that may be assigned to a Quality Control Engineer. It is possible that two or more lists may be needed to satisfy all the areas of Quality Engineering.

With such definition, Management can see how much "meat" is in the job. Employers may use it to evaluate their own Quality Engineering assignments, and individuals may use it to set up guideposts for self evaluation and improvement.

2. Committee on Professional Relations with Colleges

Thus far, the colleges offering Quality Control have done so on a limited and uncoordinated basis to those who happen to ask for it. It is the aim of the Committee to help create an organized, recommended curriculum embodying not only statistics, but also those areas of business management, principles of measurement, and physical sciences which bear upon the duties of the Quality Control Engineer. Recognition as a professional Quality Control Engineer will not come until the engineer is so trained in the colleges and universities. Curricula will not be set up for this training until a demand is created in industry for professionally trained Quality Control Engineers.

3. Management Relations Committee

The function of this committee really starts with the important and difficult task of obtaining management's general acceptance of the Quality Engineering Profession in order that a demand might be created in all industry for the Quality Control Engineer. It is only by doing this that the goal of a proper supply of adequately trained Quality Control Engineers can ever really be obtained. As Dr. Feigenbaum states,

> "I know of no single incident in product development where supply created its own demand and, as a matter of fact, economists say it never does."

A part of this acceptance will come about by coordination of the standardized work elements with industry, so that when a job applicant lists his previous experience as "Quality Control Engineer", the prospective employer will know that he has performed duties within a limited frame of reference.

A second approach to obtaining such acceptance is by demonstration of the Quality Control Engineer's contributions to industrial and public welfare through improved product quality, conservation of resources, services to public and government groups, and services to industrial standards groups. One of the hallmarks of any professional group is its ability to render organized and direct service to the public. The following are some suggested ways in which the Society can demonstrate its professional qualities:

a. Quality Improvement of the U.S. Manufactured Product

Due to many causes, the general quality level has dropped in recent years to such a degree that a national effort for quality improvement is suggested. Needless to say, it would indeed be a public service for professional Quality Control Engineers to take an active part in such activities. One approach that could be taken in this effort is for A.S.Q.C. to help in the development of product quality standards where another organization is more directly responsible for the actual setting of standards, and to urge or pursue development when the responsibility for doing so is not quite clear. In situations such as these, A.S.Q.C. should not take a back seat, but should pursue these problems vigorously until adequate quality standards exist for the great majority of U.S. manufactured products. I recognize the fact that many organizations now are doing portions of this work, but for the most part, they are working in narrow areas for answers

to specific needs. A.S.Q.C. should work to see that coverage is complete and that standards represent producer and user alike. In some instances, we may be accused of usurping authority, but being the only national society devoted exclusively to Quality Control, A.S.Q.C. can logically make only one decision in this matter: we must be an official, or unofficial, referee and see to it that procedures leading to standards are fair and factual - regardless of who uses them.

b. Conservation of Natural Resources

The Quality Control Engineer can assist in the conservation of resources by several different ways. The first and most obvious way is the reduction of waste as a byproduct of better quality. In addition, there is the possibility of his technical assistance in the determination of suitability of substitutes or new materials, and in the discovery of more efficient methods. Also, there is the direct approach to the problem where conservation of material is the primary aim. For example, the supply of tungsten is limited and must be conserved independently of cost considerations. The laws of probability statistics are certainly applicable to the discovery of the economic and legal conditions which govern the uses of tungsten, and it appears to me that this could be a valid area for Quality Engineering efforts.

c. Consulting Services to Local, State, and the Federal Government

There are many potential applications for Quality Control and statistical methods in all levels of government. Only a few have been exploited, but these could be used to advertise the possibilities. To give you a few ideas where A.S.Q.C. could be of service in this area, I would first like to expand the territory to include such organizations as the Red Cross, the Heart Fund, churches, schools, and hospitals. I might cite a few instances from my own experience.

One of our local hospitals has a problem which exists for the medical profession in general - that of accurate determination of a patient's blood count. This is important because transfusions are administered primarily on the basis of a single determination which is admittedly so variable that physicians must often use a great deal of judgment instead of facts of the case. An adequate statistical experiment would undoubtedly bring about considerable improvement in the techniques and decision criteria. At this time, we are arranging to have our Quality Control Engineers assist the hospital in the designing of such an experiment.

In cooperation with one of our local judges, we have just completed a study of juvenile delinquency data in order to depict trends of delinquency in various sections of Cincinnati. This shows where it is increasing or decreasing significantly. We hope that these facts will help to determine where additional attention is needed, and where to look for ideas leading to general improvement.

On a national scope, it is my feeling that A.S.Q.C. should volunteer technical assistance when problems arise such as the one which occurred with the introduction of the Salk vaccine for mass use. At that time, we preferred to take a back seat because the problem was "too controversial" and too much in the public eye. I consider this a real failure to grasp an opportunity to do a public service and also, to gain some good publicity for the Society.

The Committee for Professional Development recommends that each Section set up a Public Service Committee to implement such activity. I should like to extend this recommendation to include each individual industry - that it make its people available to local organizations, coordinating through local Sections when possible, and to the National Society - that it look for possible ways to contribute, and that it offer A.S.Q.C. services whenever the opportunity arises.

d. Advisory Service to Organizations Responsible for Setting Standards

This is an area that has received considerable discussion and it is generally agreed that A.S.Q.C. has an opportunity for service here, although a study is needed before we really know what steps should be taken.

4. Committee on Professional Licensing

In order to obtain professional licensing for the Quality Control Engineer, State Licensing Boards must be convinced that this is a valid segment of professional effort large and important enough to warrant specific license. In some states, this is unimportant, but in others, such licensing effects a measurable increase in professional stature.

The proposed Committee on Professional Licensing will determine - through State Boards, The Engineers Council for Professional Development, and the National Council of State Boards of Engineering Examiners - the requirements for licensing in various states. Certain Quality Control Engineers will submit applications for licensing as test cases; first in New Jersey, Illinois, and Texas. Since formal training by universities is a part of the licensing requirement, this work will be closely coordinated with the development of Quality Control Engineering curricula in schools.

5. Committee on Professional Development Planning and Section Liaison

The basic aim of this committee will be to plan and coordinate the overall program for development of professionalism. Recognizing that an important factor in professional development is the Section program, a separate group within this committee will be charged with Section liaison. This committee will work with Section committees guiding them in assessing their current professional level and in preparing education and training programs to lift the present membership to the level needed for fulfillment of the Quality Control Engineering function. Also, formation of local professional chapters may be encouraged to provide programs of interest to important professional groups within or between Sections.

The Professional Development Organization is just beginning its activities, but even now it is obvious that there is great potential benefit to be derived for the science of Quality Control and for the individual Quality Control Engineer. As its work gets underway and starts to take effect, we expect to find a greater and greater demand for the Quality Control Engineer's services in all industry. Ultimately, we expect to see the term, Quality Control Engineer descriptive of the man of professional capability who is considered a necessary part of any industrial organization.

V. RECOGNITION OF INDIVIDUAL ACCOMPLISHMENT

The fifth area in which A.S.Q.C. benefits the Quality Control Engineer is its recognition of individual accomplishment. Publication of a paper in INDUSTRIAL QUALITY CONTROL or TECHNOMETRICS, or presentation of a paper at a conference or convention, is in itself recognition of the author as a contributor to industrial progress. Participation in Society activities, committee or management positions, mark an individual accomplishment. In addition to these, there are several formal Society awards which are highly prized. Typical among these are the Shewhart and Brumbaugh awards. There are also various Section and Division awards for outstanding papers or other Quality Control contributions. These may not seem too important, but any of them marks the difference between a run-of-the-mill engineer and one who is capable of independent thought and action. Present or prospective employers recognize this, of course, and either tacitly or openly make it a part of their evaluation of a Quality Control Engineer.

Most of the foregoing discussion concerns things that A.S.Q.C. is doing now for its membership. I have also presented a few thoughts regarding what the committees are proposing to do in the immediate future. The Society would also like to have the benefit of your thinking, and so I ask that you, too, exercise your imagination and pass your ideas on to the Professional Representative in your Section. He can refer it to the National Committee. You may not have such a representative now, but your Section will be asked to appoint one soon. In this way, you will have assurance that the Society can continue to develop and improve its service to the Quality Control Engineering Profession.

Before concluding, I would like to present a few more ideas for the future:

The work of the Committee on Professional Development is intended, in part, to provide industry with a means which will enable it to recognize a Quality Control Engineer when it sees one. However, this work will probably not come to full fruition for several years. In the meantime, it seems to me that a desirable step is to provide a means for separating pseudo quality engineers from the real ones. This could be done by providing some sort of certification service conducted by, or for, industries. It could take the form of standardized tests, or a set of standardized questions from which tests could be extracted. The evaluation could be made by A.S.Q.C., if desired, or by the industry itself if it had the qualified people to do it. Perhaps an interviewing service could be part of this activity.

Another service to industry which could be supplied by A.S.Q.C. might be the preparation of self-evaluation material by which industries could compare their Quality Control Organization against some sort of standard, or we could even supply a service to evaluate facilities and systems for industry. For one thing, such a program would force us to be specific about what is good and what is bad in a Quality Control Program -- a worthy objective even if approval and certification of good facilities is not the goal. True, this would entail a great deal of labor on someone's part, but I believe industry would support such an endeavor.

A service that is sorely needed is the preparation of technical information in good format for teaching and reference purposes. We have no A.S.Q.C. Manual of basic statistics, no statistics handbooks, no handbook of measurement principles and instruments. These are so badly needed that industries are preparing their own. In some cases, Sections and Divisions have been going ahead and preparing them, not waiting for the National Organization to "get off the dime". Our reactivated Education and Training Committee under the leadership of Dr. David S. Chambers, is gathering information which should eventually lead to such publications - provided they are included among the basic goals of this Committee's activities. I think that they must be included if our service in this area is to be considered adequate. It will take time and money for this, but surely these can be provided if the right sources are tapped.

Still another thing that needs to be done is to arrive at some logical definition of the relationships between quality, performance, reliability, maintainability, and all the other attributes of a product that are a part of "full customer satisfaction". There is considerable dispute centered around the question,

"Does reliability belong in quality control or engineering, or is it separate from both ?"

This happens to be the big confusion of the moment, and could just as easily be related to maintainability which is regarded by some as a part of reliability, or performance which is regarded by some as a part of quality. My own definition of Quality is broad enough to include all of these, but this is not necessarily agreed to by others.

A.S.Q.C. needs to investigate this situation from all aspects, decide in detail what the term Quality encompasses, and insure that aid is extended to all the areas that are included in the agreed-to scope of Quality. This could be a real service and could help to solve the basic confusion that exists today. It would also build into Quality Control the scope of activity that should be covered.

In summary, then, A.S.Q.C. has been highly instrumental in the development of Quality Control as a science, and of Quality Control Engineering as a profession. We have done more than anyone else in this field, and we are currently providing a wide range of valuable services to both industry and this profession. More than this, we are very actively extending these services in directions which, we believe, will provide even greater benefit to them, and, in turn, will benefit the Society itself.

But we cannot be self-satisfied ! There are still many things that need to be done -- some that haven't even been thought of yet. The scope of interest must be as broad as industry and as detailed as the finest measurement. The Society must search diligently for more ways and means to be of service, and pursue vigorously its development, in order that our profession and organization may enjoy its rightful place among the best and most respected technical professions.

MANAGEMENT OF CONVAIR-ASTRONAUTICS RELIABILITY PROGRAMS

H. F. Eppenstein, Chief of Reliability
Convair (Astronautics) Division
General Dynamics Corporation

The formulation of many current Convair-Astronautics management-level reliability policies was begun as early as January 1953 under the auspices of the Atlas ICBM Program in the Convair-San Diego Division. A general outline of a Reliability Program was completed during a series of meetings of the Atlas Reliability Committee, formed earlier in that year, and put into effect in July 1953. The committee at that time included in its membership the group engineers or their representatives, as "reliability coordinators," from all principal groups in the Atlas Program.

Later in 1953 it was decided by the Atlas Reliability Committee that, "There should be a direct organized program for determining, improving and maintaining reliability." Further, it was decided that a reliability "working group" was needed to work exclusively in planning, organizing and implementing an Atlas Reliability Program under the specific direction of the Atlas Reliability Committee.

Today, the Atlas Program and the associated WS-117L (Project Sentry) Booster and Azusa Tracking System Programs are carried on in a separate Convair division . . . Convair-Astronautics. Each of these efforts now has a firmly established reliability program being conducted by an effectively staffed and organized Reliability Engineering Organization under the direction of a Chief of Reliability. This organization is technically responsible directly to a Division Reliability Policy Committee, chairmanned by the division Chief Engineer, and including in its membership all concerned top-level division executives. Thus, during the years since 1953, a firm Convair-Astronautics management philosophy on reliability and the required implementing policies have been established on a division-wide basis covering in considerable detail present reliability practices, procedures and programs. Specific organizational elements have also been created within the division structure to implement these policies.

Convair-Astronautics Reliability Philosophy

Convair-Astronautics management today is very much aware of the complex reliability problems associated with developing more advanced-type missile weapon systems, and unmanned and manned satellites and space vehicles. This keen appreciation of such problems stems from the fact that the men who today make up the Convair-Astronautics management cadre have for over seven years had to deal on a day by day basis with a multitude of complex Atlas ICBM Program reliability problems. It is no accident, therefore, that the present management reliability philosophy reflects the circumstances under which this philosophy came into existence. From the outset, sheer necessity demanded a reliability program that emphasized <u>direct action</u> and <u>control</u> as primary, immediate considerations. Associated supportive activities, in particular continuous research into reliability theory and analytical and statistical methods development, were carried on throughout the program, but not as the dominating theme of the effort. Even though a number of the Atlas Weapon System basic design concepts were decided upon prior to the introduction

of an organized reliability program, Convair-Astronautics management recognized early in the research and development stage that reliability considerations both "before" and "after the fact" were essential to an adequate reliability program. Present Convair-Astronautics management reliability philosophy therefore encompasses the full extent of these broad, underlying principles as well as the necessary management support to implement them . . . from the conceptual stage in operations research through predesign, design, development, production and continuing throughout operational use of a product.

Reliability programs, to be worthwhile, must be primarily "working level" programs. This "applied" type philosophy can be summarized in a single statement—detect weaknesses as early as possible and effect rapid corrective action to bring about constant improvement in product reliability. Such weaknesses may be in concepts; in equipment designs; in test procedures; in parts and materials selection, procurement and use; in production practices; in inspection methods; or in man-machine relationships from the design of equipment in engineering through the conduct of flight tests at the Test Bases. All efforts are primarily directed toward attaining the highest possible product reliability by the time the product is put to operational use.

Although aimed specifically at product operational reliability, both "before" and "after the fact" actions are essential parts of an adequate reliability program. For example, Convair-Astronautics Reliability Engineering Organization personnel in the "before the fact" situation review designs and design changes, perform statistical design of experiments to aid in specific design tasks, aid in performance of tolerance and circuit design analyses, assist in the selection of parts and materials for specific design applications, control parts standards in design, and are responsible for the reliability orientation of design and development engineers. In the "after the fact" case, personnel from the organization actively work with all concerned departments from engineering through production and product support—as well as being assigned and residing at Test Bases where they work with reliability problems and program activities, however, the basic program philosophy emphasizes the "action" approach to reliability improvement over the "theoretical" approach.

In accord with and implementing this basic division philosophy, division management reliability policy is specific on the following points:

1. Recognition of the need for the constant improvement of product reliability and achieving the utmost degree of cooperation to this end are direct "line" responsibilities of all division management.

2. A distinct, planned and organized reliability program, aimed at realistic accomplishments and scheduled in accordance with the product development program to which it applies, is an essential element in all programs involving product development undertaken by the division.

3. All division reliability programs will be continuously monitored by an active Division Reliability Policy Committee made up of top level division executives who will assemble frequently and regularly to consider reliability policies and problem areas.

4. Division reliability programs shall be the direct responsibility of a Reliability Engineering Organization which will have as its primary assignment the work associated with the implementation of these reliability programs.

5. The Convair-Astronautics Reliability Engineering Organization shall be directed by, as his primary responsibility, a Chief of Reliability . . . a member of the division management team . . . who reports technically to and is a member of the Division Reliability Policy Committee.

6. The Reliability Engineering Organization shall be deliberately structured on an adaptable basis with provisions made for the addition or deletion and/or modification of groups and functions with a minimum of disturbance . . . this practice to allow efficient adjustments to be made when the necessity arises for handling one or more reliability programs simultaneously with parent programs involving widely varying types of products such as missile systems, interplanetary probes, and manned satellites.

7. Reliability Engineering Organization staffing shall be maintained at a high technical level, including scientists or engineers with specialized training and experience in the conduct of reliability programs as well as in one of the various fields of engineering such as electronics, mechanics, physics, or human engineering.

Reliability as a Management Responsibility

Recognizing that reliability considerations can only enter into all division operations in which they should be involved if accepted by all levels of supervision, Convair-Astronautics management has made reliability program participation a required "line" responsibility throughout the division. This policy emphasizes the early and continued indoctrination of all management personnel with reliability program elements and procedures. Supervision, from first-level supervisors through all levels up to the division manager, are directed to acquaint themselves with reliability principles and the factors in all operations under their control that affect product reliability. It is an expressly stated responsibility that they shall actively concern themselves with promoting reliability improvement in their areas.

Also, under the authority of existing Division Standard Practices (DSP's), they are instructed to actively cooperate with the Reliability Engineering Organization in the reliability indoctrination of their personnel, both from a general knowledge point of view and as such indoctrination applies to day by day tasks in their areas. In addition to the preceding, a variety of special tasks, specifically related to division reliability activities, are required of various personnel throughout the division's structure. Examples are such functions as failure reporting, operating problems reporting, and the exercise of individual, personal ingenuity and initiative in recommending changes or improved methods or procedures that contribute to improved product reliability. The latter is reserved as a final alternative in bringing such suggestions up for management consideration to be used only after unsuccessful use of routine channels. In summary, reliability, as a vital aspect in all division activities, is a direct responsibility of each member of the management

team as well as a part of the normal working activities of every employee of the division.

Reliability in the Product Growth Cycle

The product growth cycle at Convair-Astronautics, Figure 1, begins in operations research--where systems are originated and/or studied, on as quantitative a basis as possible, for suitability, and, in conjunction with the Reliability Engineering Organization, for their initial and ultimate reliability. These studies include system compatibility with respect to national capability--cost of initial procurement and later upkeep in terms of money, manpower, materials, machines, and management--operational concepts required in their use--and the number of years required to attain the desired operational capability or effectiveness.

As these broad weapon system characteristics are studied--configurations emerge that warrant closer scrutiny and comparative evaluations. Ultimately--a single approach is selected--often with several alternative variations--with reliability considerations among the dominant factors in this selection. Such decisions usually rest with management or the customer or both and are made on the basis of the results obtained in the various analyses and studies.

The process now becomes predominantly a problem in the pre-design area. At this point the problems that must follow if a contract is awarded--in functional diagramming--in subsequent detailed design work--in parts and materials selection and procurement--in providing necessary production facilities--in manpower--in management--and in the need for considering in detail the actual reliability problems involved in producing and using the new product begin to have their effect. Here then--in the planning for new products--anticipated maintenance--probable operating procedures--expected demands in various types of personnel training--and a host of other reliability problem areas enter the picture. Since most of these considerations are critically important to the end use reliability of the product, the Reliability Engineering Organization has specific responsibility for inserting inputs into the pre-design effort. Inputs from a multitude of other engineering groups are obtained through operations analysis, pre-design studies, and during later general and detailed design efforts. Contributing engineers, among other things, specifically concern themselves with the reliability aspects of such problems as new product application, specifications and test requirements; materials handling controls; new production methods and techniques; and essential production process controls and facilities. In general, an increasingly close liaison is maintained between the various parts of the organization throughout the later or more detailed design stages. This close coordination then continues after completion of the design and during the development, production, test, and operational use stages of product development programs.

Reliability Program Functional Elements

Convair-Astronautics management, in recognizing the critical need for an effective, planned and organized reliability program associated with each product development undertaken, also anticipated a probable need for certain variations in these programs. For example, the reliability problems to be dealt with in an ICBM weapon system as compared to

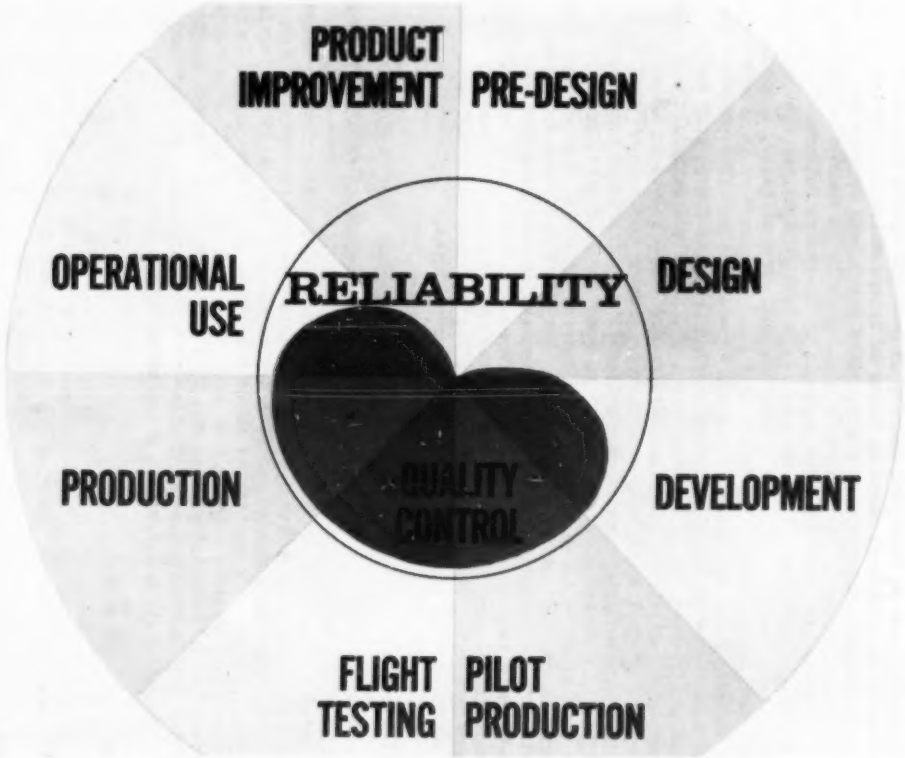


Figure 1. Reliability and Quality Control in the Product Growth Cycle

a manned satellite, although having much in common, would necessarily also include some significantly different problems. Human errors, for instance, would affect the ICBM Weapon System only during design, development, production, testing and up to launching of the individual missiles. In the case of manned satellites, the limitations of human capabilities under conditions of close confinement, high accelerations, free fall, and possible elevated temperatures after vehicle launching would undoubtedly have an important bearing on over-all system reliability. This would be true even where no means of intervention into system operation by the occupants were possible, since provisions made to protect and maintain the lives of the occupants would greatly increase the complexity of the system and therefore the probability of system failure. As would be expected, therefore, the emphasis on certain reliability problem areas would be shifted from program to program and, in some cases, elements would need to be added to programs that are not necessary in, for example, the current Atlas Reliability Program. Present reliability programs are implemented in the Reliability Engineering Organization in four basic areas . . . research . . . operations . . . parts and materials . . . and laboratory facilities. A reliability program best suited to a manned satellite project might well include much greater emphasis on such elements as human factors and system reliability, which would then become relatively independent areas and at the same level organizationally as the four elements established for the current reliability programs. Personnel to provide initial staffing for these activity areas are already part of the Reliability Engineering Organization--assigned to these functions within the various groups. Thus, consistent with Convair-Astronautics management policy for reliability, the Reliability Engineering Organization is structured for rapid, efficient adaptation and expansion to handle a wide range of reliability program tasks.

Division Reliability Policy Committee

Essential as a policy formulating body representing top-level division management, the Division Reliability Policy Committee, Figure 2, is an integral part of all division reliability programs. The existence of such a group serves as a constant reminder to all division personnel of the cardinal importance management attaches to division reliability activities. Further, when this committee is an active group, as is the case in Convair-Astronautics, meeting regularly to deal with reliability policy problems, corrective and remedial actions can be effected directly through one or more of its members--thus originating with top management and carrying all the authority necessary to bring about prompt action.

Committee Chairmanship

Initially, in 1953, the chairmanship of the Division Reliability Policy Committee--then known as the Atlas Reliability Committee--was vested in the Chief Project Engineer. Subsequently, following formation of the separate Convair-Astronautics Division, the Division Manager became chairman of the committee. At his direction, active chairmanship was vested in his senior technical assistant, the Technical Director for the division. Recently, because of the urgent need for the services of the Technical Director at Convair Division level, the duties of the chairmanship were undertaken by the Division Chief Engineer. Thus, very few changes have been made in the committee chairmanship since its inception, and all changes have been made consistent with Convair-Astronautics policy which

Figure 2. Convair-Astronautics Division Reliability Policy Committee

requires that the Division Reliability Policy Committee chairman be at all times a top-level division executive.

Committee Membership

Since its earliest existence, the Division Reliability Policy Committee has numbered among its membership the Division Manager and management level representatives from all concerned departments in the division. Prior to establishment of a separate division, each area of the Atlas Project was represented by a reliability coordinator—the group at that time being designated the Atlas Reliability Committee. A quick review of current Reliability Policy Committee members clearly indicates that all major areas . . . engineering, long range planning, contracts, production, quality control, procurement, tooling and product support . . . are represented on the committee by their respective managers.

Frequency of Committee Meetings

In accordance with a firm policy established by the Convair-Astronautics Division Manager, regular meetings have been held for the past several years once each month with few exceptions. In addition, a number of special meetings have been called to consider and render decisions on urgent policy-level reliability problems. The regular meetings of the Division Reliability Policy Committee are held on the third Thursday of each month. Special meetings may be called at any time by the chairman of the committee.

Implementation of Committee Decisions

Division Reliability Policy Committee policy decisions are primarily implemented through direct action within the committee. Decisions are set forth in the minutes of each meeting under the "Conclusions" section, and the minutes, signed by the chairman, are distributed promptly to all committee members. In a number of instances, committee decisions initiate the preparation of new or modification of existing division or departmental standard practices documents. These, after initial preparation or modification, are usually reviewed by the committee before issuance. On some occasions further study has been warranted with committee decision being withheld until after hearing one or more special reports on a reliability problem area. Such reports are frequently the responsibility of a sub-committee comprised of several members of the Division Reliability Policy Committee and appointed by the chairman.

Reliability Program Monitoring

A routine function of the Division Reliability Policy Committee is the monitoring of division reliability programs. This is accomplished through regularly hearing various reports on reliability program activities and the review of charted or published data provided by the Reliability Engineering Organization. Data provided in chart form, covering various types of progress on the over-all system, are reviewed in the and development phase as gross indicators of system trends.

Reliability Engineering Organization

The Reliability Engineering Organization, Figure 3, is administra-

tively a portion of the Development Section of the Division Engineering Department but is technically responsible directly to the Division Reliability Policy Committee. Organization operations and the conduct of division reliability programs are responsibilities of the Chief of Reliability as head of the Reliability Engineering Organization.

Chief of Reliability

A member of the Division Reliability Policy Committee, to whom he is responsible technically, the Chief of Reliability has as his primary assignment the direction of the various functions of the Reliability Engineering Organization and the conduct of all division reliability programs. He is directly charged with carrying out the policy decisions of the Division Reliability Policy Committee and with the adequacy and effectiveness of reliability activities throughout the division. As a specialist in the field, the Chief of Reliability represents the division in reliability matters involving customer relations as well as with interested government agencies and industry. He is responsible for acquiring the necessary skilled and experienced technical personnel and the facilities to assure an efficient organization and effective reliability programming.

Organizational Structure

Consistent with over-all division management reliability philosophy, the structure of the Reliability Engineering Organization is maintained on as adaptable a basis as is possible. For example, a change in the organization soon to go into effect is illustrated by comparing Figure 3 with Figure 4, the latter organizational configuration shows a markedly increased emphasis on the human engineering aspects of several new reliability programs being initiated. The organization is currently divided into four groups, with each group composed of three distinct but somewhat related functions. These functions are the lowest-level organizational elements involved and are each under the direction of a senior engineer specializing in a particular reliability activity. Each group is under the direction of an Engineering Group Supervisor. The present organizational structure is well suited to the conduct of the various reliability programs since it insures a balanced application of reliability program effort in all the necessary areas. However, as established, the organization can be quickly reoriented to deal with different program variables or expanded to undertake additional programs as required.

Personnel Requirements

Consistent with division management policy, personnel assigned to the Reliability Engineering Organization are, for the most part, trained engineers experienced in reliability work. Of the total of 64 persons in the Reliability Engineering Organization, 45 are qualified design, research or electronic engineers; six are test engineers, associate engineers or engineering or technical assistants; two are engineering writers; and seven are research and development technicians and engineering aides. Of the engineers assigned, three are design specialists and 27 are rated as senior engineers in their respective classifications. Educational background in the group includes six doctorate degrees, two masters degrees and 35 bachelor degrees in various engineering fields, physics, and human engineering (psychology).

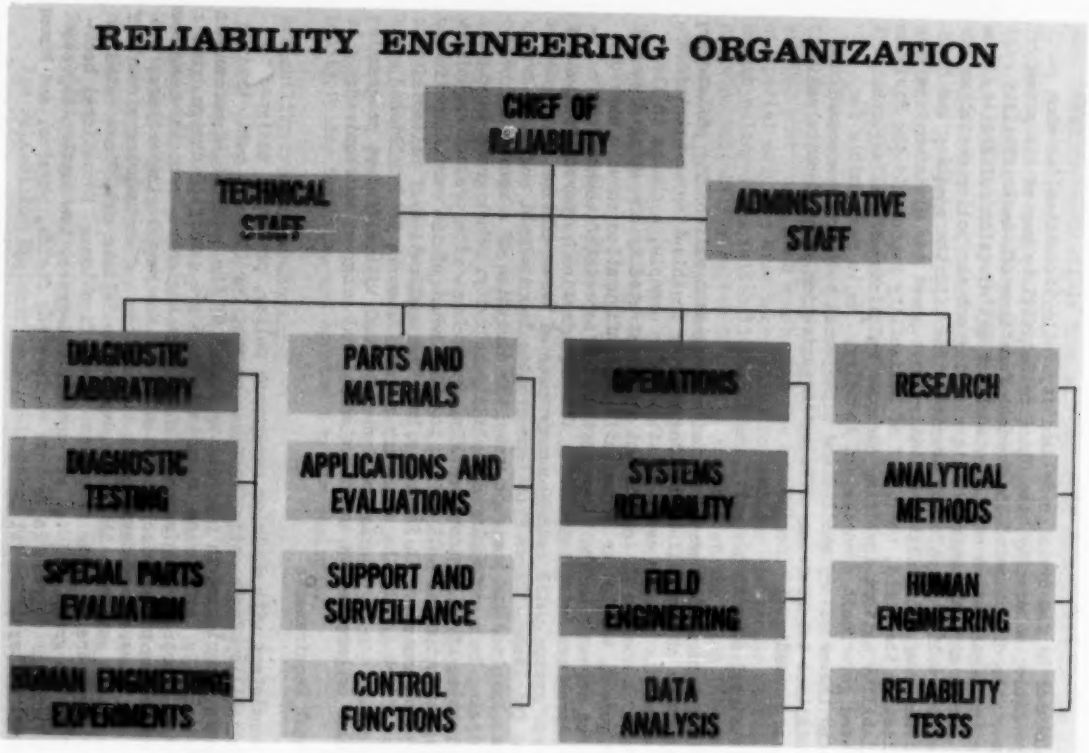


Figure 3. Present Reliability Engineering Organization Structure

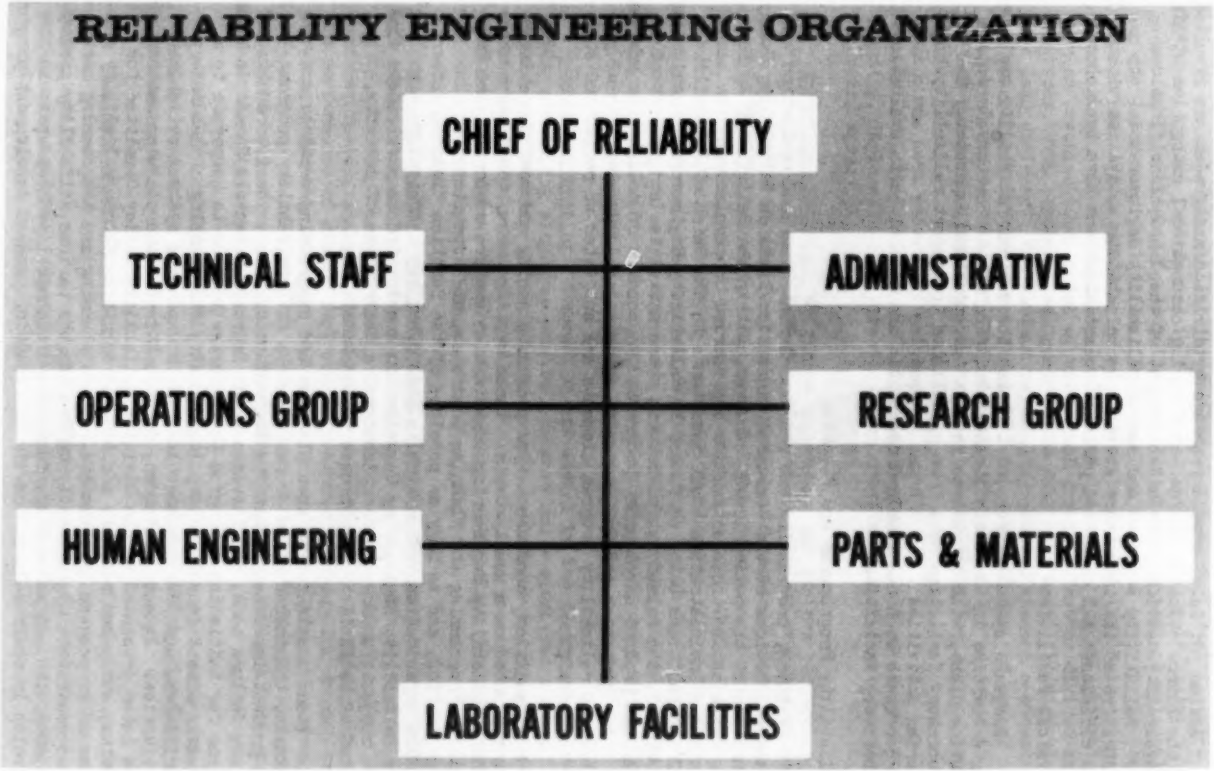


Figure 4. Scheduled New Reliability Engineering Organization Structure

Action initiated over a year ago by division management has resulted in the introduction of two new engineering occupational categories within Convair-Astronautics. The establishment of these new classifications . . . reliability engineer and senior reliability engineer . . . are indicative of the importance attached to reliability activities within the division. All qualified personnel classifications have been changed.

Physical Facilities

Adequate space is provided to house the various reliability engineering activities in the administration building of the new Convair-Astronautics plant, with provisions made for organization expansion. Two laboratory facilities have been established to support reliability engineering activities . . . a human engineering laboratory area and a much larger space to house combined diagnostic and special parts evaluations laboratory facilities.

Reliability Programming and Planning

Planning, programming and program element study and evaluation are primary functions of the Convair-Astronautics Division Reliability Policy Committee at the policy level. However, within the Reliability Engineering Organization these activities are carried on regularly by the technical and administrative staff assistants to the Chief of Reliability. All the functions being performed by Reliability Engineering Organization personnel, the relationships between these activities and conduct of over-all product development programs, the conduct of reliability programs at other companies and in government agencies, and investigation of possible new reliability program elements are subjected to review and study by this staff.

In addition, conference-type program planning and review meetings are conducted on a weekly basis. These meetings are attended by members of the Reliability Engineering Organization who are members of the Reliability Programming and Planning Committee, all supervision in the organization, and as many as possible of the function "lead" engineers in the organization. From time to time other key technical personnel in the Reliability Engineering Organization are invited to participate when the subjects under consideration involve their areas of activity. The Chief of Reliability is chairman of the committee.

Items considered in such planning and programming meetings are, for example, the introduction of new elements into reliability programs--such as product-improvement training in production areas, the introduction of a new failure code, the issuance of an improved Materials Standards Manual, plans for initiating a program to encourage vendors to cooperate to a greater extent in reliability improvement, the planning of laboratory facilities, and problems associated with expanding systems reliability activities. In many instances, the results of these planning meetings are later presented to the Division Reliability Policy Committee for review, study and possible adoption as solutions to new reliability problems or as program improvements. These planning and programming meetings also contribute significantly to keeping the various reliability programs in step with the scheduling, funding and technical problem areas as they develop in their related Convair-Astronautics product development programs.

At present, the reliability programs being conducted at Convair-Astronautics are implemented in four broad areas: (a) reliability research, (b) reliability operations, (c) parts and materials reliability and (d) laboratory facilities. Each of these areas of activity in the program have a corresponding group within the Reliability Engineering Organization. To more closely control and integrate program activities from an organizational point of view, each group is further divided into functions, with personnel assigned to the functions under the direction of a lead engineer. The various reliability program element assignments are established at the "function" level, but are each part of one of the integrated over-all reliability programs. The activities assigned to the various groups and functions of the entire Reliability Engineering Organization are:

RELIABILITY RESEARCH: Reliability research activities include the planning, implementation, direction, and coordination of functions in the following areas:

Analytical Methods: Originate approaches and perform researches leading to the development of advanced mathematical and statistical methods applicable to measuring and predicting, if possible, various reliability indices; calculate tolerances for and perform error analyses on parts, components, subsystems and systems; apply most advanced mathematical and statistical methods to the design and support of experiments; perform significant analyses and evaluations of search for critical weaknesses testing and other data; originate and formulate suitable advanced training courses in related areas of reliability measurement methods for engineering personnel, and conduct these courses.

Human Engineering: Consult on and assist in the design of control and/or monitoring displays; perform research and evaluations of man-machine relationships in equipment designs; review, study, and analyze for evaluative purposes all levels of operations and procedures such as countdown procedures. Develop and implement reliability program indoctrination and training liaison activities, Figure 5.

Reliability "Search-for-Critical-Weakness" Tests: Develop, schedule and implement search for critical weakness testing; perform research leading to extending the usefulness of such tests in determining reliability factors; review, analyze and evaluate the results of such tests; direct the performance of statistical and mathematical evaluations of such test results; coordinate findings with other Reliability Organization groups and functions as well as with other Convair-Astronautics Departments; recommend corrective actions on the basis of test results and initiate corrective actions.

PARTS AND MATERIALS RELIABILITY: Reliability program parts and materials activities include the planning, implementation, direction and coordination of separate functions in the following areas:

Application and Evaluation: Direct and coordinate specific applications research and evaluations on mechanical, hydraulic, pneumatic, electro-mechanical, and electronic parts as well as on broad classes of engineering materials; assist design engineers by providing specialized technical support in areas of parts and materials applications.

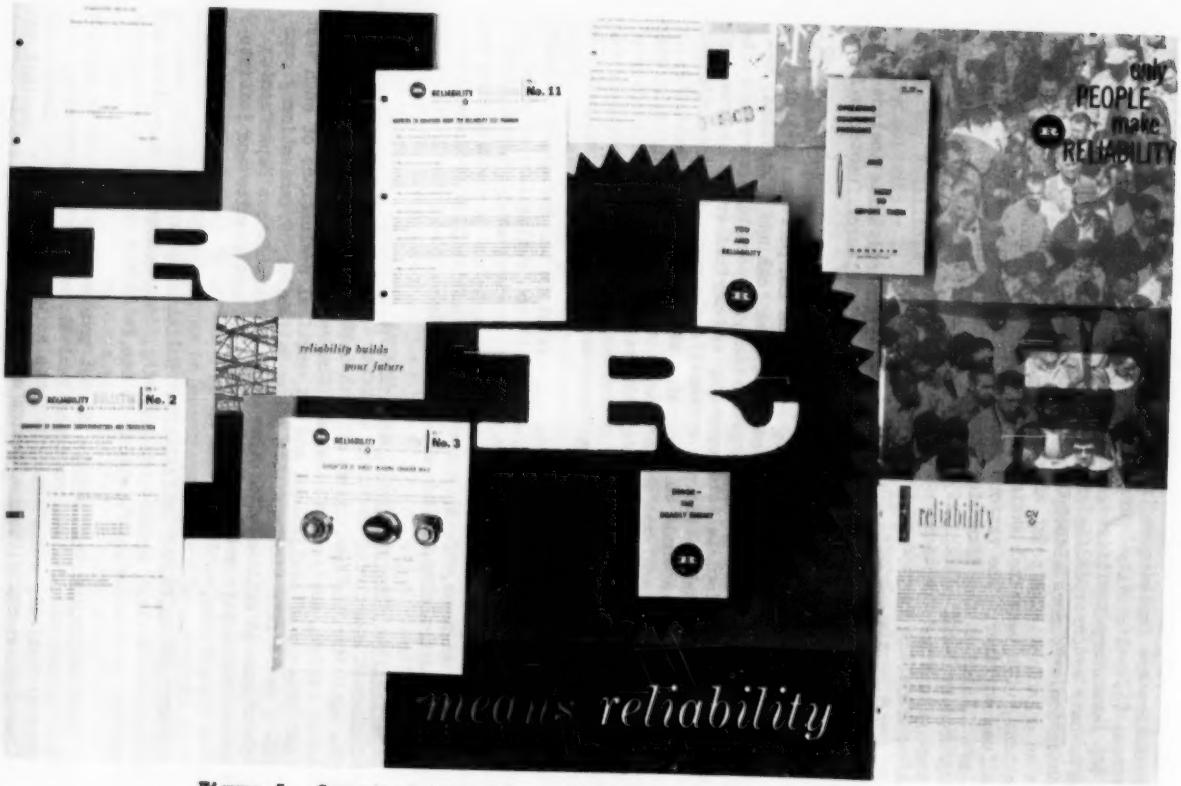


Figure 5. Convair-Astronautics Reliability Indoctrination Materials

Liaison Support and Surveillance: Perform general factory area parts and materials "use" liaison and surveillance and maintain close liaison with the Quality Control Department; review and analyze outgoing parts and materials documents; support Aircraft Industries Associated (AIA) activities and perform as Convair-Astronautics AIA representative in areas of parts and materials reliability.

Control Functions: Originate or review and analyze for approval all requested Standards for Convair-Astronautics parts and materials requirements; review parts and materials requirements; review parts and materials specifications and tests; exercise design control of specific parts and materials applications; originate, publish and disseminate Reliability Test Requirements (RTR) used by receiving inspection personnel in the Quality Control Department as the basis for accepting or rejecting incoming parts and materials.

RELIABILITY OPERATIONS: Reliability program operations activities include the planning, implementation, direction and coordination of separate functions in the following areas:

Systems Reliability: Review, study, analyze and prepare evaluations on equipment components, subsystems and systems for reliability improvement; recommend improvement actions and follow-up on such recommendations; review specifications and control drawings; perform consultative and surveillance activities in general production and field use problem areas; provide chairmen for Malfunction Diagnosis Teams in accordance with established procedures; initiate, plan and perform analyses and evaluations of systems interactions and implement corrective actions; initiate and/or perform and report on failure diagnoses, malfunction diagnoses, or failure analyses.

Field Engineering: Maintain cognizance over the various reliability programs field activities, including those at Test and Operational Bases; carry out follow-up on actions requested by Reliability Engineering Organization personnel in the field and on actions to be taken by these personnel as directed by the organization; insure timely compliance on the part of Reliability Engineering Organization personnel in the field with contractual requirements.

Data Analysis: Initiate and implement methods for acquiring, interpreting, and processing into semi-permanent records—including electronic accounting machine tabulations—failure, consumption, operational life, and other data classed generally as reliability data; perform research into and conduct regular analyses of methods, techniques, practices and procedures for improving the effectiveness, accuracy, and intelligibility of collected data, and to present these data in various ways such as to be most effective as a corrective or informative feedback into the respective reliability programs; satisfy, in a timely manner, all contractual requirements.

RELIABILITY LABORATORY FACILITIES: Reliability Laboratory facilities have been planned to serve the needs of all engineers in the organization assigned to reliability problems that warrant laboratory work. Staffing of the facilities with strictly laboratory personnel will be kept to a minimum, thus encouraging participation in laboratory activities from time to time by all Reliability Engineering Organization personnel.

Facilities provided in the human engineering, diagnostic and special parts testing laboratory areas supply test data to support much of the Reliability Engineering Organization activity involving specialized laboratory effort. However, to avoid duplication, extensive use is made of other division testing facilities and, in some instances, of independent commercial laboratory capabilities where such capabilities involve extensive or large scale testing requirements.

Diagnostic Laboratory: Diagnosis of failed parts and assemblies, determination of parts failure modes, and special parts evaluation techniques developed from this work.

Special Parts Evaluation Laboratory: Special tests to establish Reliability Test Requirements (RTR) for parts and materials in support of engineering and production activities on a division-wide basis; evaluation of new parts and materials to special requirements for improved reliability; statistical testing of parts and materials to provide data of the type needed for mathematical analyses.

Human Engineering Laboratory: Reliability human engineering design test facility; evaluation of display-control parts and components; man-machine research on operational procedures and countdowns; research into "man into space" problems.

Diagnostic and special parts evaluation laboratory facilities are capable of performing MIL Standard tests on most electronic parts essential to the various programs. Special equipment for high precision work in known reliability problem areas is also developed as a routine activity in conjunction with state-of-the-art advancements.

Several principal equipments, aimed at specific areas of testing, are of interest.

X-ray: The X-ray facility is planned to provide the most advanced type high-definition, constant-potential X-ray known to be available today for the examination of electronic and electro-mechanical parts. Low voltage equipment for plastics and non-ferrous thin-walled sections, and high voltage for heavier electro-mechanical parts are included. The total voltage range available is from 18 to 300 KV, using three X-ray tubes. A fluoroscopic X-ray image intensifier will be used for direct visual examination of parts. This intensifier-X-ray system constitutes what amounts to a significant advancement in the state of the art and makes possible an important refinement in failure analysis through the use of high speed motion picture photography of sealed, operating parts. One of the most advanced-type high speed motion picture cameras available today will be used, capable of picture speeds in excess of 6,000 per second, and a rapid film developer and motion analysis projector are planned to support this work. This equipment will also work with a vibration generator so that X-ray motion picture studies of sealed and enclosed parts can be made while they are operated in an environment involving controlled vibration levels.

Infra-Red: Infra-red measuring equipment is available which will provide color-photographic and visual presentations of infra-red images. Associated with this are a film reading densitometer which will provide means for measuring the exposure of film records and, to augment the basic

equipment, an evaporograph and densitometer with automatic timing and film advance used to increase the accuracy of test results.

AN INDUSTRIAL ENGINEER LOOKS AT STATISTICAL CONTROLS

James L. Lundy
Special Projects Supervisor
Josten Manufacturing Company

What is Control?

Control is one of the most widely misunderstood functions of management. It often is considered synonymous with accounting or the combination of accounting and budgeting. However, neither accounting nor budgeting constitutes control, for the first is simply a data-gathering function and the second is simply a technique for planning and evaluating performance. Since control involves action of a guiding or restraining nature, the submission of a plan and the reporting of data must be followed by guiding or restraining action before control occurs.

To discover a theft, production lag, or sales decrease is only to be cognizant that a problem exists. To take remedial action for a specific case and preventive action against future cases which are similar in nature is to control.

The above statements are not made in an effort to belittle those of you in accounting or budgeting. Instead they are intended to provide the proper perspective for our discussion. To say that accounting or budgeting is contributory to control rather than the essence of it should in no way imply that either function is unimportant.

By way of illustration, we might liken accounting to a thermometer which indicates the temperature in a particular structure. Even though it is the thermostat that is the control device, the role of the temperature-measuring device is vitally important. If the information-providing thermometer does not perform properly, the thermostat's control will be less than satisfactory.

When discussing managerial control I invariably think of a birth control program which failed not long ago in India, for the Indian situation constitutes a good correlative to the inadequacy of control found so commonly in our respective firms. It seems that the women of India were shown a birth control system based on the use of a string of beads, the beads being white except for the three center ones, which were red. The explanation given with the beads was that each day a bead was to be moved from left to right, and on any day when a white bead was moved, abstinance would not be necessary. However, as is sometimes the case with our business controls, the system was neither understood nor applied correctly. Unfortunately, the women involved in the birth-control study believed that the white beads cast a magic spell, so whenever a red bead indicated abstinance contrary to their wishes, they simply moved enough beads to get to the magic white bead area and proceeded on their merry way.

In the business world, as in India, control systems are not always understood. As a result, we get distorted data, too much or too little data, information which is too late to be of value, and data which are not really pertinent to our projects, to mention only a few of our control problems.

Why Relate Control to Other Functions?

At the outset of an undertaking management should establish its goals, or objectives. It should also decide on some means for recognizing the attainment of these objectives, or, in other words, it should establish some criteria of success. Next, systems of reporting should be devised in terms of the criteria selected. Effective control then--and only then--can be effected through the interpretation and comparison of progress and plans.

I cannot stress too strongly the importance of preceding control with clear definitions of goals and authority-responsibility relationships. If you don't have a goal, how can you possibly believe in control? In other words, if you don't know where you want to go, how can you stay on a path to get there? Similarly, if you don't have clearly defined authority and responsibility, who are you trying to control? In other words, who should be correcting things and who should be improving his performance?

Your reaction to these comments might be that if you're going to have perfect organization and planning, there will be no need for control, so why have it at all. I must admit agreement that in the face of perfect organization and planning time devoted to control would be wasted, for it is imperfection in people and things that gives rise to the need for control. However, we must keep in mind that perfection does not exist, and in view of this fact the importance of control must not be slighted.

What About "Statistical Controls"?

Having discussed the nature of control and the interrelationship of control with other functionsof management, we are now in a position to examine the concept of so-called statistical controls. However, before we proceed we should first note that in view of the nature of control already described, we must consider the term "statistical controls" as being a misnomer of the first degree. Control is accomplished by people, not numbers. To say "statistical controls" makes no better sense than to say "stastical planning", "statistical organization", and so on. At their best, statistics are simply indicators of control needs. At their worst, they can lead to the ruination of your firm.

In taking a Cook's tour of the area loosely called statistical control, perhaps a logical starting point would be a re-evaluation of the concept of scientific management and our use of operating data. Although I consider myself an avid proponent of scientific management, I am extremely fearful of the consequences of what might be called psuedo-scientific management.

I believe that it is possible to become so enthusiastic about statistics that their use can actually be detrimental, rather than beneficial, to the decision-making process. We have all met individuals who feel that life without statistics would hardly be worth living and that the most appropriate way to evaluate a report is to consider the quantity of statistical data contained therein. As Robert H. Roy notes in his book, The Administrative Process.

"Numbers tend inordinately to dominate decision making by pushing aside intangibles which may exceed in importance that which is measureable and by acquiring an aura of accuracy. Operation gaming, Monte Carlo methods, linear programming, time study, accounting, statistical quality control, and aptitude testing are all examples of quantification of attributes which are essential to decision making. They should not be neglected; however, when they are deified they are detrimental to sound decisions."

This worshipping of numbers is leading, unfortunately, to relatively widespread substitution of mathematics and statistics for courses of value in various professional fields. Although I am strongly in favor of students in industrial engineering and the various divisions of management having solid appreciation of mathematics and statistics, I am against substituting so much mathematical and statistical work for other valuable courses that the identity of the original professional group is almost lost.

I want to be sure that you don't misunderstand the above comments. Mathematics and statistics are wonderful fields and have contributed significantly to progress throughout the world. My point is that managers, engineers, doctors, and other professional men should become well acquainted with mathematical and statistical techniques, but should rely on other specialists--namely, mathematicians and statisticians--for guidance and detailed mathematical and statistical work. I am apprehensive that if we continue to move in the direction we are heading, the substitution of mathematics and statistics for other subject matter will lead ultimately to the loss of distinction between any two professional groups. In my own field, industrial engineering, I have seen considerable evidence that this process has gone far beyond a healthy point already.

One amusing aspect of this situation is the "sophistication jag" currently popular in so many areas. Many universities are "sophisticating" such fields as industrial engineering and business management by substituting "sophisticated" theoretical course work (e.g. mathematics and statistics) for applied courses. Each time I hear the term "sophisticated" used so proudly by someone in this regard, I wonder to myself is he realizes that Webster defines sophisticated as "Not in its pure, natural, or original state; adulterated; amended unwarrantedly; deprived of native or original simplicity; made artificial or highly complicated."

Along with the worshipping of quantification which I have been discussing is the equally important problem of lack of technical soundness in the application of mathematical and statistical techniques. Whereas the first problem is traceable to a narrowness in outlook and imbalance in development, the second is traceable to incompetence or human fallibility. Perhaps it is not too far afield to say that some individuals treat the existence of quantified information as a personal invitation not to think. Let me cite several of the situations or cases upon which these comments are based.

One nationally known--and supposedly progressive--firm has used the ratio of indirect labor to direct labor as one of its major control

indices. Throughout the company pressure has continually been exerted to maintain this ratio below a certain critical level--even, unfortunately, at the expense of higher unit and total costs and lower profits! It is interesting to note that one of the fastest and easiest ways to achieve a lower ratio of indirect to direct labor is to hire more direct labor personnel!

In educational circles we find similar situations, even though our great universities are supposed to be leading us toward continually improved technology. Actually, my limited experience in the field of teaching has led me to believe that many of our universities are, administratively speaking, still in the dark ages. Among their more dubious practices is that of basing faculty requirements largely on a ratio of faculty to students. One of the consequences, of course, is that even in those cases where large classes are infinitely practical, class size is held down so that there will be no trouble justifying large--or larger--faculties.

In another industrial case, we find a progressive corporate giant with a "wonderful" management development program. How do they know their program is so good? The answer is easy to find. They simply show you the fact that of those people having had the benefit of the management development program, 90 percent have had a promotion within the past two years, while among those executives not having benefited from the program, only about 25 percent have won promotions during the same period. The fly in the ointment appears, however, when a little detective work reveals that (a) only the most superior men are selected for the management development program, and (b) when an opening occurs, it is filled preferably by someone who has had the "benefit" of the program. It doesn't take very high-powered thinking to bring out the fallacies involved here, and to indicate that, as far as this company can tell from the data used, its management development program may be doing more harm than good.

In another instance we find a firm using its automatic data processing equipment twenty-four hours a day, seven days a week. The result is stacks upon stacks of reports loaded with information. Unfortunately, the firm's organization is so loose and confused that people either don't care what the information signifies or they don't know how to take action based on the data. Another difficulty is that the data themselves are known to be inaccurate because of incomplete or inaccurate information input, so nobody places any faith in the reports. The net result, as you might guess, is almost complete lack of control.

On a more limited scale, we can cite the use of statistical control charts by inspectors using worn, broken, or dirty gages. In each of these instances we have found situations where so-called statistical controls were either improper in themselves or were improperly applied.

Again we should pause to note that we cannot issue a blanket indictment against statistical controls, but must simply criticize their being overworked or improperly established and applied.

One of the common mistakes in the treatment and use of quantified information is a failure to establish--or even realize--the conditions or assumptions upon which the particular analysis must be based.

So often we use analytic techniques based on the condition of randomness without knowledge of this limiting condition and/or without paying any attention to the satisfaction of the requirement of randomness.

A similar situation exists in the use of formulas for comparing and analyzing economic expenditures. How many of you who have used these formulas realize that they are invariably based on uniform series applications? Although these formulas offer an easy way to get answers to some of your questions, care must be taken that the answers obtained are satisfactory for your decision-making needs.

The point which I would like to make regarding statistical controls is simply that they are industrial dynamite. If used properly, they can help move mountains. If misused, they can contribute to the rapid decline of your enterprise.

How Does Reporting Fit Into the Control Picture?

Reporting systems can be likened unto our nervous systems, which contribute to personal control. In both cases we need and have sensory systems which transmit information to the brain or manager.

As was mentioned briefly at the beginning of this paper, control is inherently tied in with the achievement of plans. Therefore, reporting systems which are designed to facilitate control must be geared to specific information needs. Goals which are sought must be defined, and the measures of achievement, or lack of it, must be established. The reporting system then is established as the means for providing the controlling person with the data which he needs for evaluating conditions and taking action.

You don't have control based on good data if you simply nip here and there periodically in an effort to get isolated data on first one problem and then another. Such pecking away at only one problem area at a time provides only a part of the total picture essential to sound control, and is indicative of incomplete and poorly planned reporting and control activity.

By establishing clearly defined goals and equally clear definitions of responsibility, authority, and accountability, you have laid the foundation necessary for a sound reporting framework. Having done so, you have greatly enhanced the likelihood of achieving a reporting system which gives neither too much nor too little information, neither too much nor too little accuracy, reporting periods which are neither too long nor too short, and reporting dates which are neither too early nor too late.

While we are on the subject of reports, perhaps a brief comment or two on the use of automatic data processing equipment would be appropriate. Automatic data processing systems can contribute significantly to better control by reducing clerical errors and even total information processing costs. Perhaps one of their greatest contributions is the provision of data sufficiently early to permit the actual controlling of one's business course rather than the after-the-fact regretting that so frequently occurs when slow clerical methods are employed. On the other hand, managers must be careful that the glamour

of automatic data processing does not lure them into the development of reports which are both too voluminous and numerous to be consistent with the needs of the enterprise.

Can We Have Better Control?

Most of you have your attention focused right now on this matter of control. However, if you had been caught off guard last week with a question such as "Do you have good control," many of you would have replied that you do, in fact, have "reasonably good" control. Unfortunately, the chances are you would have simply considered that you have a controller or comptroller, a large accounting group, a budgetary control program, a statistics department, and automatic data processing equipment. Although you might have mentioned a few more items (or some different ones) to support your position, I hope that by now you realize that reasons such as these are not, of themselves, indicative of sound control! The items I have just listed are simply tools which can be used in your control function. The real issue concerns not the existence of such tools, but the extent to which their utilization has contributed to the best possible operation of your enterprise.

When the question regarding the improvement of control is studied in view of the results of our control efforts rather than on the basis of our mere application of a variety of tools and techniques, I believe we must conclude that better control is not only possible--it is essential to long-term survival in our competitive business atmosphere.

Conclusion

Eventually, I hope, we will have a greater number of top quality management personnel who can establish and maintain control systems (including so-called statistical controls, where appropriate) of a positive and beneficial nature. However, at the present time I think that American business as a whole is suffering from a critical need of managers who can serve competently to organize all functions so that statisticians and other specialists provide more useful data and much less garbage.

References

Lundy, James L., Effective Industrial Management, The Macmillan Company, 1957.

Roy, Robert H., The Administrative Process, Johns Hopkins Press, 1958.

Schleh, Edward C., Successful Executive Action, Prentice Hall, Inc., 1955.

STANDARDIZATION OF QUALITY CONTROL TERMS AND SYMBOLS

Irving W. Burr, Purdue University

1. Limitation. Although this session at the National Convention has been arranged by the Standards Committee, of which I have been a member for eight years, I am here speaking as an individual person, rather than as a member of that Committee. In this way, and only in this way, can I be free to say what I wish to say.

2. Standardization of Terms. The complete job of building a set of standardized terms is a far more difficult job than the outsider would guess. There is first the problem of determining which terms to define. Few ASQC members would come up with the same list of terms to be defined. For this reason many people are consulted. Valid reasons for omitting a term can be (1) that it is already widely agreed upon and not easily misinterpreted, (2) that it is insufficiently important, (3) that it is in such widespread use under too hopelessly many meanings to be worth "rescuing". Terms included are those of importance, which it will be possible and desirable to clarify and standardize, without thereby imposing a "straight-jacket".

Once the list is agreed upon tentatively, some committee member spear-heads the job with a first draft. This is then batted back and forth by correspondence and in committee meetings until there is substantial unanimity. In a committee of five unanimity is commonly possible, although it may require some giving in on unimportant points (but not vital ones).

The third general step is to publish in Industrial Quality Control the first draft, which is open for suggestions from all ASQC members. These are collected, sorted with reference to the item in question and given careful study. Diametrically opposite viewpoints are frequently expressed on the same item. Also a great many of the suggestions have already come up in committee discussion. But all are weighed and given careful consideration.

At last a final draft is submitted to the Board of Directors of ASQC, which may or may not give approval in its precise form. When approved by the Board, the Standard can be published in final form in Industrial Quality Control and separately.

By its very nature a Standard must be a compromise, not wholly perfect in anyone's eyes. But it is an agreement which can be useful, even vital. It is like our agreement to drive down the right hand side of a road. Even though some might prefer the opposite side, we must all agree to the convention, to make any progress, elsewhere than to the grave.

3. Standardization of Symbols. Undoubtedly the most controversial problem in symbols is that of the standard deviation. We therefore shall discuss it. Letting n be the sample size and N the population size we have the following two systems in rather common use:

	ASQC	Other
Population mean	$\bar{X}' = \frac{\sum_{i=1}^{N} X_i}{N}$	$\mu = \mu_X = \frac{\sum_{i=1}^{N} X_i}{N}$
Population std. dev.	$\sigma'_X = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N}}$	$\sigma = \sigma_X = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N}}$
Sample mean	$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$	$\bar{X} = m = \frac{\sum_{i=1}^{n} X_i}{n}$
Sample std. dev.	$\sigma_X = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}}$	$s = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}}$.

The unifying principle in the ASQC system is that unprimed quantities are the sample values, while those with primes are population or standard values. In other systems the population values are given the Greek letters corresponding to the Roman letter for the sample value. Either system seems to be consistent within itself.

Now let us give especial consideration to the definition of and symbols for standard deviation. I myself prefer the definition of sample standard deviation as adopted by ASQC, namely

$$\sigma_X = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}} .$$

Note especially the n in the denominator. My reasons for preferring this definition are as follows: (1) The use of n in the denominator makes good sense to the beginner, facing a standard deviation for the first time. As a teacher I can easily "sell" him on its reasonableness as a sort of average of the deviations. But I have seldom had any success getting the n-1 definition across to him. (2) The use of n fits in with the descriptive value of the standard deviation for the data actually at hand. (This same descriptive use may be made for the deviations from a best fitting multiple correlation relation when the number of points is rather small. Division by n - k - 1 destroys all descriptive value to the data actually at hand. Significance tests, etc. can still be made by appropriate adjustments.) (3) The use of n comes directly out of many maximum likelihood calculations. (4) A most significant point in favor of the use of n in the denominator is that of the consideration of bias. If we use n as in

$$\sigma_X = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}} \ .$$

then we know that to get an unbiased estimate of the population standard deviation, we have to divide by c_2. Then

$$\sigma_X / c_2 \text{ averages } \sigma_X' \ .$$

There is also bias in estimating $\sigma_X'^2$ from $(\sigma_X)^2$. In fact

$$\sigma_X^2 \ \frac{n}{n-1} \text{ averages } \sigma_X'^2 .$$

Thus σ_X and its square are both frankly and avowedly biased, and we must do something to remove the bias. On the other hand when we use

$$s^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1} \ .$$

we have a quantity which averages exactly σ^2, the population variance. This sounds as though s^2 were therefore much better than the n definition. But the big disadvantage is that although s^2 averages σ^2, s does not average σ ! In fact to average σ we must use

$$\frac{s}{c_2} \sqrt{\frac{n-1}{n}} \text{ averages } \sigma \ .$$

This causes no end of confusion for students. Moreover an unfortunate number of textbook writers are not clear on this either!

Some persons even go so far as to say that use of n rather than n-1 is wrong. This is obviously absurd, as one can define standard deviation with any denominator he wishes and then make adjustment for any purpose he has in mind. As a matter of fact, in using Student's t, if we defined

s' to be $\sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-3}}$, then $t' = \frac{\bar{X} - \mu}{(s'/\sqrt{n})}$ would have a far faster approach to normality than does t. This is because $\sigma_{t'} = \sqrt{(n-1)/(n-3)}$, and not 1. In fact one hardly needs t' tables beyond n=10.

Finally, (5) the definition of standard deviation with n in the denominator is simpler and requires little adjustment even in the F formula. And of course (6) the definitions are standard for ASQC, and appear in Industrial Quality Control, some other journals and some textbooks, as well as in international standards organization.

Having said all I can in favor of the ASQC definitions I have nevertheless decided that I must now personally recommend that we go to the other system of notation. This is recommended by me for the following reasons: (1) Most other areas of statistics than ASQC now use the other notations. By cooperating fully there is an excellent chance that all statistical organizations can work out a standard set of notations. (2) The average level of sophistication of ASQC members is going steadily upward, despite the number of beginners coming in each year. These beginners surely ought to be able to learn one or two more Greek letters. They already must learn Σ and σ, and they seem to enjoy the latter, as witness the ASQC seal. The main new letter is μ to replace $\bar{X}'$, and is not very hard to learn. (3) Actually the use of small Greek letters instead of primes is easier, since we do not have squares to contend with such as $\sigma_{\bar{X}}^2$. (4) Those of us in teaching and/or regular contact with those in other statistical areas know that there is at least some tendency for persons to look down upon ASQC because of notations used. It seems absurd to any objective student of notations, but the tendency is present. (5) The prestige and large membership of ASQC is such that if we change, there is solid hope that a world-wide standard set of symbols can now emerge. This must be close to the optimum time when such a highly desirable result can come about by courageous and generous action on our part.

4. A Final Word. The foregoing is no sudden whim, as I have faced the problem in thirty years of teaching statistics, out of innumerable textbooks. And in fact, I went through the completed manuscript of my textbook to change it to ASQC symbols.

PROGRESS IN AIR FORCE QUALITY CONTROL

Bruce C. Dunn, Colonel, USAF
Quality Control, Hq AMC

During the past year the Air Force has been faced with the same factors which have forced and are forcing industry to reorient many of its quality control concepts and philosophies. Qualitative, rather than quantitative, superiority has become more vital than ever. Changing Air Force logistic concepts have resulted in many major changes in the Air Materiel Command's methods of doing business and these changes in turn have caused adjustments in Air Force quality control activities.

Most of the factors causing us to reorient our programs are readily apparent such as the ever increasing complexity of our weapon systems, which in turn results in more complex tests and more complicated test equipment. Today we face demands for fantastic degrees of reliability; increasing costs of weapon systems; spectacular failures which occasionally occur; demands for production compression; and finally, the paradox of trying to provide greater assurance of quality of more complex items on an ever decreasing operational budget. The interplay of these factors has caused many changes to be made and only time will tell if all of these changes can properly be classed as progress.

For purposes of this discussion, I have classified our activities over the past year into the following general areas:

First - Proficiency - What we have done to keep pace with the "state of the art" by improving the caliber of our personnel.

Second - Organization - Certain changes have been made to keep pace with the job to be done which we feel will enable us to be in a better position to cope with the day-to-day problems which arise.

Third - Technological - What we have done to improve quality control methodology, techniques and procedures.

Fourth - Management - What we have done to improve our management of the program.

Improved Proficiency

Our greatest efforts during the past year have been towards improving the proficiency and skills of personnel throughout the Air Force Quality Control organization. While this need was recognized by various Air Force agencies, it was also confirmed by a private consultant as being one of our most pressing problems. Consequently, we have taken a number of steps to introduce new skills to keep pace with technological advances and to upgrade the skills of personnel already in the organization. We believe that one of the most significant of these steps has been to introduce the use of professional engineers throughout our Air Force Quality Control organization. So far, professional personnel with backgrounds in the fields of chemistry, electronics and mathematics have been hired. Most of these people have been utilized in the missile

fields but we intend to extend the use of such personnel into all areas. To upgrade the capabilities of our assigned personnel we have established three training courses under the Air Force Institute of Technology at Wright-Patterson Air Force Base, Dayton, Ohio. These courses were developed by the Ohio State University Research Foundation in collaboration with Headquarters Air Materiel Command, with the assistance of Mr. L. A. Seder, author and consultant. Mr. Seder, who I am sure many of you know, is also monitor of these training courses and we have been fortunate to have such people as Dr. Juran; Mr. McLaughlin of General Electric Company; and Mr. Purcell of Raytheon Manufacturing Company as guest lecturers. The courses given are all three weeks in length and consist of a course in basic statistical quality control, one in advanced statistics and analysis and a general course on quality control management. They are open to both military and civilian employees of the Air Force including personnel from the operating commands as well as Air Materiel Command. In addition, a number of spaces are being made available to the other services.

To provide an input of officer personnel into our Quality Control program, we have established a limited program of graduate training in Quality Control engineering at Stanford University.

As a part of our effort to improve the caliber of our personnel, we have done considerable work in the past year to improve the Civil Service job standards under which our people are hired. To those of you in private industry, this probably seems like a simple problem; however, the standards under which our Civil Service people work have been established for some time and must, of necessity, be rather broad in scope and sometimes do not recognize the newer skills which we need to get the job done. Consequently, we have made a concerted effort during the past year to get the job standards covering our Quality Control people updated. We feel we have made considerable progress on this project.

To assist in training our technicians on the latest methods of nondestructive testing, a color film has been produced and is now available to anyone interested in that particular subject. Film No. TFI-5226, "Air Force Non Destructive Testing" is obtainable from the Air Force Film Library Center, Air Photographic and Charting Service, USAF, 8900 South Broadway, St. Louis, Missouri.

Organizational Changes

During the past year, a number of changes have been made in our Air Materiel Command Quality Control organization. As part of the overall AMC decentralization program, Quality Control operating functions are now in the hands of our Air Materiel Areas and depots with only staff functions remaining at Headquarters AMC in Dayton. To support this concept of operation we have augmented the Quality Control Office at each of our Air Materiel Areas so that each of these offices can assure a comprehensive and coordinated quality control program throughout all areas of their responsibility. These Quality Control Offices report to the Air Materiel Area Commander. To emphasize an across-the-board approach to quality control and avoid special emphasis on any particular area such as procurement, supply or maintenance, our Headquarters Air

Materiel Command Quality Control office was realigned into an Engineering Division, an Equipment Division and a Planning Division. The most significant aspect of this realignment is the establishment of the Quality Control Engineering Division. This Division was created primarily for the purpose of enhancing our engineering capability and is largely staffed with professional personnel having engineering backgrounds or degrees in mathematics and statistics.

Technological Advances

In this area a great deal of work has been done during the past year. Perhaps the most outstanding effort has been the development of a single world-wide integrated calibration system. This system, designed to assure the accuracy and compatibility of our weapon systems throughout the world, is essentially a closed loop system providing accurate calibration of basic standards and test equipment. Formerly much of our equipment was returned to the depots for calibration. Under the new program some 168 Air Force Bases throughout the world are being equipped with basic sets of standards which will enable them to more effectively calibrate their equipment. These bases will service satellite bases and facilities in their immediate areas.

In addition to the work of improving our internal Air Force system, a survey of the calibration systems of our contractors has been conducted in three phases. The first phase consisted of a two-part questionnaire prepared jointly by Aircraft Industries Association, National Bureau of Standards and the Air Force and was circulated to member companies of Aircraft Industries Association, Electronic Industries Association, and National Security Industries Association. The second phase consisted of a physical survey of the calibration systems of Air Force contractors. These surveys were conducted by our Air Materiel Areas and the final phase was a visit to selected ballistic missiles contractors by a team composed of representatives from Air Materiel Command, Air Research Development Command and the National Bureau of Standards. The purpose of these surveys was to isolate problem areas in the industry calibration system and assure that adequate and positive calibration is being accomplished.

In the statistical area, a great amount of work has been done both by our own personnel and by contract. We have participated with the other services in a number of research contracts. These contracts cover such studies as a "Model for the Control and Investigation of Complex Items" and an "Investigation of the Use of Standard Statistical Quality Control Techniques with the IRAN Concept." In addition, AMC Manual 74-23 "Multi-Level Continuous Sampling by Attributes" has been converted into a Department of Defense Handbook (H-106). MIL-STD-414, "Sampling Procedures and Tables for Inspection by Variables for Percent Defective," has been issued and is being implemented throughout our Quality Control activities, and I am sure many of you know MIL-STD-105 has been revised. Considerable other developmental work such as a handbook on "Reliability and Life Test Sampling" and a technique for "Single Level Continuous Attribute Sampling Plans" is in progress.

No doubt one of the major accomplishments of the past year has been the publication of a joint military specification on Quality

Control which will supersede our present Air Force Specification, MIL-Q-5923. We, along with the other two services, have actively supported the publication of this document and feel that it is a major step forward in assuring equitable treatment of all contractors and in providing optimum assurance of quality at minimum cost.

I have not touched on the subject of reliability although considerable work and study have been accomplished on this subject by various Air Force agencies, as this will be covered by Colonel Griffith in a separate presentation.

Management Improvements

Our main effort in this area in addition to improving the proficiency of our management has been to establish a system for gathering and analyzing management data. As you all know, the problem of determining the effectiveness of a quality control program is, at best, a difficult one. We feel that this job is even more difficult in the case of a military quality control operation because we do not have the dollar sign attached to many of our operations as industry does. However, we are working in each of the major areas of supply, maintenance and procurement to establish a system which will provide us with information regarding the overall effectiveness of our program, the general quality level of materiel at the various points in the logistic cycle, the costs of operating our quality control program, and some workload measures for use in establishing better manning standards. In addition, in the maintenance area we are attempting to establish work standards for quality control which will measure the efficiency of the operation more effectively.

One of our biggest jobs in the management area just as in industry is that of selling the quality control function to personnel throughout the command. Mr. David Hill of Hughes Aircraft Company has outlined this problem most explicitly in his pamphlet "Communicating Quality Control Ideas." If you have not seen this document I recommend it to you. In addition to the usual program of providing quality control releases to appropriate internal publications, the publication of a Quality Control bulletin and providing quality control posters and bulletins for use throughout the organization, we are presently in the process of producing a short color film describing the Air Force Quality Control program. This film will be used in orienting Air Force personnel with respect to the quality control function. Our emphasis has been on establishing an understanding that quality control is a management tool which will provide management with the best information it needs to make decisions with regard to one of its basic responsibilities - the responsibility for quality.

In summary, I would say that our most significant improvements during the past year have been the introduction of engineers into our Quality Control organization, the strengthening of our Quality Control organizations in each Air Materiel Area, the contributions we have made to the development of statistical methodology, and, finally, the actions we have taken to improve our management of the overall Air Force Quality Control effort.

QUALITY CONTROL AND RELIABILITY OF MILITARY EQUIPMENT

John J. Crowley, Vice President
Grand Central Rocket Co.

As you are aware, the theme to be developed by this panel discussion is "How Quality Control Can Help Improve Reliability of Military Equipment." From the standpoint of our National Defense, it should be clear to everyone present that the achievement of the highest possible reliability is not just desirable, it is an absolute "must." I cannot speak for the other industries represented here, but for the Missile Industry, the targets are getting smaller and more distant. This is, of course, not true literally -- it only seems that way. It is another way of saying that military procurement requirements on allowable miss distance are becoming more demanding every day. Since the American position is basically a defensive posture, our Armed Forces must know, with a high order of confidence, that if it is necessary to press the firing button, the missiles launched will accomplish their mission. In case of national emergency, almost is not good enough. There is only a pass-fail criterion to be applied in this situation. The consequences of even a few failures can be national disaster. There can be no room for unreliability in our field.

In one of my previous papers, I developed the thesis that reliability requirements may have an effect on an organization. In that paper[1], it was concluded that there was a tremendous organizational impact which could be met in a number of different ways. At my present company -- Grand Central Rocket -- the reliability function was conceived, organized and integrated into the Quality Division. Many of my comments on how quality control can aid reliability will be drawn directly from our own past experience.

In order to intelligently discuss today's question, I feel it is first necessary that we define the term "reliability." What is reliability? Today, this word has almost as many interpretations as interpreters. However, reliability must be recognized for what it is -- a design parameter ranking in importance with weight, size, performance, etc. From this definition, it is axiomatic that reliability cannot be produced or inspected into an item during production or in the field. It must be introduced at the basic design level. Recognizing, therefore, that the inherent reliability of a product must be established in the design stages, it is readily apparent that control of quality, or quality control, must begin at the same time, by early review of design. But we cannot stop here; product reliability must be actively pursued and continuously improved and monitored throughout the life cycle of the product. Therefore, once the design reliability has been achieved, it is the classic responsibility of the Quality organization to monitor the manufacturing process to help guard against degradation of the reliability from this source. Further, unless the product is designed for completely interchangeable usage, it is also a responsibility of the Quality Division to help prevent misapplication of the product through customer liaison and field investigation. Therefore, the modern quality

[1]The Impact of Reliability on Organization in the Manufacture of Airborne Electronic Equipment -- J. J. Crowley

control system must be organized and staffed to aid in improving reliability by vigorous activity in the areas of design, manufacturing, and operational use.

How does the Quality Division operate in these three areas? I will discuss each of these separately and briefly. First, consider the area of design. Obviously, this is a highly specialized area. The primary responsibility for developing the original concepts of the design must lie with the design engineers. Reducing ideas to a set of design parameters is the important function of this group. However, once the details of implementation are completed on paper, questioning should begin. It must be remembered that many times, in producing a detailed design, a large number of parameters are "traded off." For example, if the safety factor of a case is high and performance low, the designer may trade a little weight for additional performance. Too frequently in the past, designers have been wont to trade reliability for additional performance. Too often, reliability is reduced through unnecessary complexity in the preliminary design. As you are probably aware, one of the major characteristics of solid propellant systems is the simplicity of design that is possible. While there are unavoidable complexities with sub-assemblies, basically the solid rocket is:

1. A solid propellant grain.
2. A suitable container for the grain.
3. A device to ignite the propellant.
4. A suitable opening to vent the exhaust gases.

At our company, to assure reliability and maintain maximum simplicity, all designs are reviewed by both Engineering and the Quality Division. Engineering concerns itself primarily with design features, technical requirements, feasibility, and concept excellence. On the other hand, Quality reviews designs for redundancy, reproducibility, allowances for all expected environments, inspectability, presence of quality requirements, adequate safety factors, and use of proven components and concepts. All groups concern themselves with the question, "Is this the simplest and most reliable design to do the job?" Unless all requirements are realized, the necessary changes must be made here. It should also be remembered that cost-consciousness is important to the Quality Division. Paper re-designs are much less expensive than changes to an existing product.

Secondly, having aided in making reliability an inherent part of the design, it must be remembered that at this point, all we have is a "paper product." The bulk of the Quality job has been carried by the two technical sections -- the Reliability Group and the Quality Engineering Group. From this point on, the full force of the Quality Division must be brought to bear. As stated earlier, after design review, the role of the Quality organization now expands to that of maintaining the inherent design reliability -- by monitoring and improving the "manufacturing reliability factor." Vendor selection, sampling plans, test plans, inspection plans, data collection, failure analysis and reporting, etc., are the tools used by Quality Engineering, Assurance, and Reliability to maintain over-all system reliability. Table 1 has been prepared to indicate how the many functions performed aid in maintaining reliability.

Thirdly, having assisted in the design concept and having assured that the product meets the designer's requirements through production, one would think that the Quality Division would be ready to rest. This is not the case! Operational usage in the field can affect the reliability as much as either design or manufacturing. For example, suppose an individual unit has been proven to be 95% reliable; now if the unit is used singly as intended, then it is expected that five out of 100 units may fail. However, if the application requires the simultaneous use of two of these units, as in hypersonic sleds, then the probability of success for the combination would be expected to be 90.75%. Thus, it is readily apparent that operational use is important. The Quality organization must therefore assist in preventing misapplication of the end item through customer and field liaison.

To summarize briefly, the goal of reliability -- PERFORMANCE WHEN YOU NEED IT -- can be achieved only by logical and well-conceived design effort, effective application of operational data to product improvement, and an awareness of the danger of degradation throughout production and operational usage. Therefore, it is up to each of us to extend this total concept into our organizations so that Quality Division personnel can take an active and more dynamic part in improving reliability of the product.

Table 1

FUNCTIONS PERFORMED BY QUALITY CONTROL TO IMPROVE RELIABILITY

Responsible Quality Control Section	Function Performed	Reliability Improvement
Reliability - Data Section	Vendor rating	Vendor selection on basis of: (1) Product excellence (2) System excellence (3) Delivery (a) Past performance (b) Present requirements
Reliability - Experiment Design Section	Component test design planning	1. Component testing, proofing strength levels 2. Statistical analysis of stress conditions 3. Provide data for re-design
Reliability - Experiment Design Section	Planning statistical test designs for motor development program	1. Statistical analysis of environmental and development changes on ballistics 2. Predictions of reliability growth rate
Reliability - Reliability Engineering	Reliability co-ordination	1. Co-ordination of all reliability effort
Q. C. Engineering - Laboratory Section	Process control test development	1. Manufacturing reliability maintenance

Table 1
(continued)

Responsible Quality Control Section	Function Performed	Reliability Improvement
Q. C. Engineering - Laboratory Section	Material testing	1. Manufacturing reliability maintenance
Q. C. Engineering - Laboratory Section	Material specification	1. Manufacturing reliability improvement and maintenance
Q. C. Engineering - Engineering Section	Process and equipment capability studies	1. Manufacturing reliability improvement
Q. C. Inspection - Receiving Section	Inspection of vendored parts	1. Assurance that vendor items meet designer's requirements 2. Failure reporting
Q. C. Inspection - Process Section	Inspection of propellant processing	1. Assurance that manufacturing meets designer's requirements 2. Failure reporting
Q. C. Engineering - Laboratory Section	Process control	1. Assurance that propellant meets designer's requirements 2. Failure reporting

Table 1

(continued)

Responsible Quality Control Section	Function Performed	Reliability Improvement
Reliability - Test Inspection	Inspection of ballistic testing	1. Assurance that propellant testing is done per design and reliability requirements 2. Failure reporting
Reliability - Data Section	Data collection	1. Summary of vendor, process and test data
Reliability - Reliability Engineering	Data analysis and reporting	1. Analysis and interpretation of experiment data 2. Analysis of failure reports
Reliability - Reliability Engineering	Corrective action	1. Manufacturing reliability maintenance
Reliability - Reliability Engineering	Training	1. Maintaining reliability awareness among all personnel
Reliability - Assurance Engineering Section	Q. C. auditing	1. Auditing of all Q. C. inspection and engineering activity to maintain high performance

Table 1

(continued)

Responsible Quality Control Section	Function Performed	Reliability Improvement
Reliability - Data Section	Management reports	1. Program process reports to management
Reliability - Reliability Engineering Q. C. Engineering - Engineering Section	Field investigation	1. Determination of the use and effectiveness of the end product

CONTINUOUS FORECAST CORRECTION IN PLANNING PRODUCTION IN THE APPAREL INDUSTRY

by

ROBERT E. HEILAND
Research Director
KURT SALMON ASSOCIATES, INC.
Consulting Management Engineers
Washington, New York, Nashville, Greensboro, Atlanta

In a large number of industries today, the combination of a seasonal sales peak and a changing line of products makes the problem of balancing inventory and production seem a constant headache. Certainly one of the most severe hazards of business life is the risk of guessing wrong concerning market trends. This also poses a severe restriction on attempts to utilize ever the simpler forms of mathematical programming in these industries.

The research described in this paper is an attempt to take a new look at the opportunity to forecast successfully, and to keep a finger on the accuracy of such a forecast, through the use of statistical control methods. While the application described here is drawn from a portion of the men's and boys' apparel industry, the same method would undoubtedly be beneficial in any industry selling consumer's goods, where the line of products must be designed in advance of a selling season, and must have breadth and variety of appeal to the market. The carpet, toy, hardware, electric appliance, housewares, shoe, jewelry, office equipment, and furniture industries, among many others, could find use for this type of attack.

In a large portion of the men's and boys' clothing industry, a manufacturer's "line" is created two to three months before any garments are sold to retailers. While many factors are involved in creating a line, the fabrics and styles do not change completely each season for most firms. There are some "staples" and some new items, along with a group of items which are either on their way up to greater consumer acceptance, or are one their way down to either the status of a staple, or to oblivion.

Decision as to what styles and fabrics will make up a line is only part of the planning problem. Each may be offered in 4 to 8 colors, and in 6 to 15 sizes. One maker of boys' dress trousers had 10 styles, 22 fabrics, each in an average of 6 colors and 6 sizes, making a total of 7,920 items to plan, each being one inventory item to control!

The problem of acquiring piece goods from the mills as they are needed, but in large enough quantities to get a reasonable purchasing economy; a 2 to 4 week manufacturing lead time; and economic quantity limitations in cutting and sewing, all impose severe restrictions on planning.

Taken together, these factors would not be insuperable if the total market demand for the season could be known in advance of manufacturing. But production must begin before any goods are sold, and working capital requirements throughout the industry permit virtually no end-of-season inventory of finished goods. Exhibit 1 shows a representative seasonal pattern of orders received, production, and shipments.

This research was not directed toward a total solution, but was an attempt to neutralize two factors - color and size - in the planning of production, over the major portion of the line, by providing statistically reliable means of determining and evaluating shifts in demand patterns within a season and between seasons.

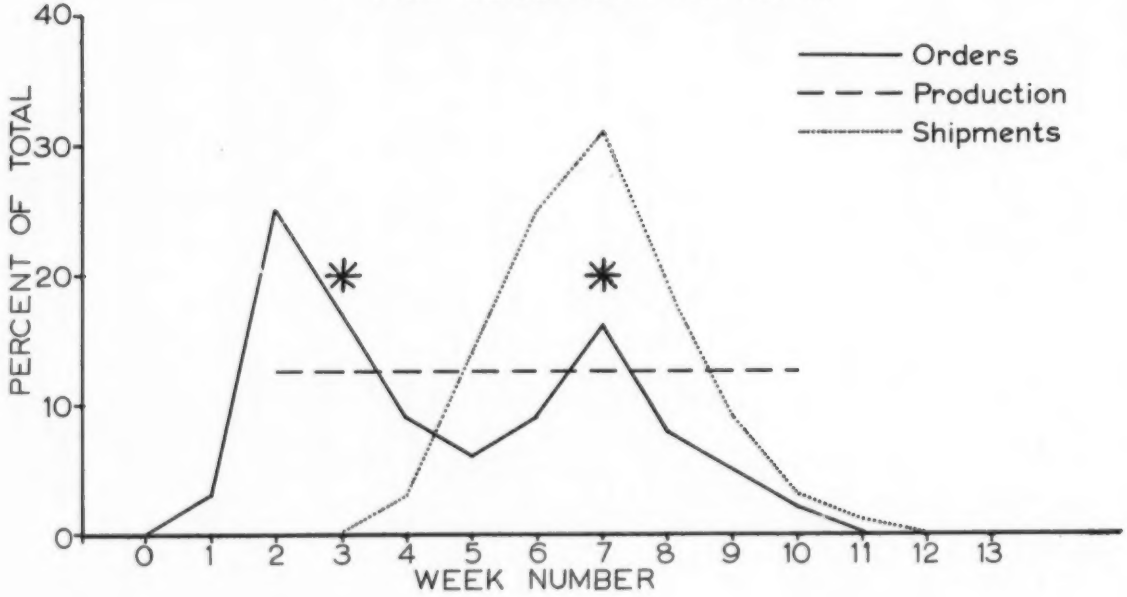
Planning of how to cut piece goods to achieve the right distribution of colors and sizes involves some knowledge of the cutting process used in nearly all garment manufacture. A bolt of piece goods is rolled out on a long table, first in one direction, then in the other, so that it piles up in layers; a pattern "mark" is then made, usually on paper laid on top of the pile of fabric. The entire pile is then cut according to this pattern mark, and the stacks of pieces are taken off and enter production, to be sewn into garments. In making a spread, or pile of fabrics, different colors of the same fabric may be spread on top of each other (although they are taken off and processed separately), but the pattern mark determines the distribution of sizes which will result from the manufacturing. Thus, a pattern mark may result in 4 sizes of trousers in the proportion of 15%, 30%, 35%, and 20%, throughout all colors and fabrics making up the lay. Usually, no permanent mark is used, but the mark is made up each time a fabric is spread, in accordance with the distribution desired by management.

Since a line must be filled out to meet shipment dates during the season, there are usually several or many cuts of the same fabric and style during the production season. In advance of the season, as well as during the height of the ordering period, the precise distribution of sizes and colors which the customers will want is not known. As a result, the manufacturer is faced with three alternative ways of arriving at a cut size distribution, or color distribution:

1. Constant adjustment of the size or color distribution to agree with the latest order total to date.
2. Cutting throughout the season to the same scale, with the conviction that all orders may be met safely during the season.

Exhibit 1

THE SEASONAL PATTERN



3. Cutting according to a pre-set size scale distribution early in the season, adjusting this distribution during the season only when evidence indicates that a significant chift in the distribution has occurred.

Most manufacturers use the first method, even though their facts may be in error until late in the season. Sometimes the error factor results in significant amounts of substitutions, "seconds" and dead stock, while other sizes and colors wanted by customers have not been made in a quantity sufficient to meet demand.

There is a need to find a scheme for standardizing cuts in terms of the distribution among sizes and colors within a style offering. This would require development of a device to "flag" items in which continued use of the standard cut distribution is leading to unbalanced inventory, and/or inability to ship satisfactorily, and/or undesirable amounts of substitutions.

The economy in cutting to the same, or nearly the same, distribution throughout a season lies primarily in 4 areas:

1. Ability to effect more standardization in the cutting room operations, by saving layouts for re-use, thereby avoiding some layout costs.
2. Ability to effect more economical piece goods utilization through minimizing waste allowances; better piece goods acquisition planning; and maximizing working capital turnover.
3. Salesmen in the field can offer this standard distribution to customers, when solicited to do so, as the most probable distribution of sales. This assists the customer to avoid bad inventory unbalance in his own stockroom.
4. Above all, to permit better management planning. This comes about by permitting the firm's executive staff to concentrate more of its thinking on the newer, less stable items, by requiring less attention on the high-volume, stable items.

Item 4 recognizes not only the "exception principle" but also the "A-B-C Inventory Control" principle. This states that a company's line of product usually breaks rather sharply into 3 segments (using illustrative figures only):

A. 10% of the items in the line account for 50% of volume;
B. 40% " " " " " " " " 40% " " ;
C. 50% " " " " " " " " 10% " " .

The common practice of treating all itmes in the line with equal attention and time for planning, using the same procedures, fails to recognize the true relationships of these items. In this industry, most profit accrues through careful attention to the "C" items. Devoting more management time to the "C" items by providing more mechanical methods for planning the "A" items assists in applying the exception principle.

The following hypothesis represents the heart of this new approach:

"A Law of Large Numbers distribution of the sizes of boys exists. Therefore, there is a basic distribution of sizes of trousers to fit these boys. However, the basic distribution is truncated and re-truncated by several industry practices; among these are the use of size ranges for regulars, longs, huskies, and little boys; use of both staple and newer, high-style items; and marketing in several types of distribution outlets.

"Within each eventual distribution, the probability that a boy will wear one or another size is clearly a reflection of the proportionate volume which his size is of the total for all sizes offered in the distribution.

"Therefore, if the basic distribution volume characteristics for each group of items in the line can be established, based upon experience, a size scale distribution can be adopted as standard, which will be correct. While this basic distribution holds true, any variations in composition of orders with respect to size-volume relation, will be a function of the probability limits imposed by the Binomial expansion. Consequently, small samples may show more variation from the true distribution, proportionately, than may large samples."

Similarly, market acceptance or preference for colors in which a style is offered, is equally amenable to this type of thinking.

At the start of a season, as production of staple items is begun, use is made of the historic average distribution of sales by size and color within a given style and fabric. As the season progresses, two or three checkpoints are set up for evaluating the statistical significance of data at hand. At these times, the order distributions received to date are compared with the limits imposed by the Binomial, and the existence of truly significant shifts can be isolated readily. For the situation shown in Exhibit 1, the check points are set up at the times shown by asterisks, when approximately 45% and 85% of orders are in.

Exhibits 2 and 3 are plotted from actual data in one company using this method. Each of these exhibits shows experience over one full season, plus data available at the first checkpoint, in the following season. The Binomial limits are shown as the symmetrical envelope on each side of the centerline established by the preceding season's records. These control limits are obtained by the conventional approximating formula:

$$\bar{p} \pm 3 \sqrt{\frac{\bar{p}(1-\bar{p}}{n}}$$

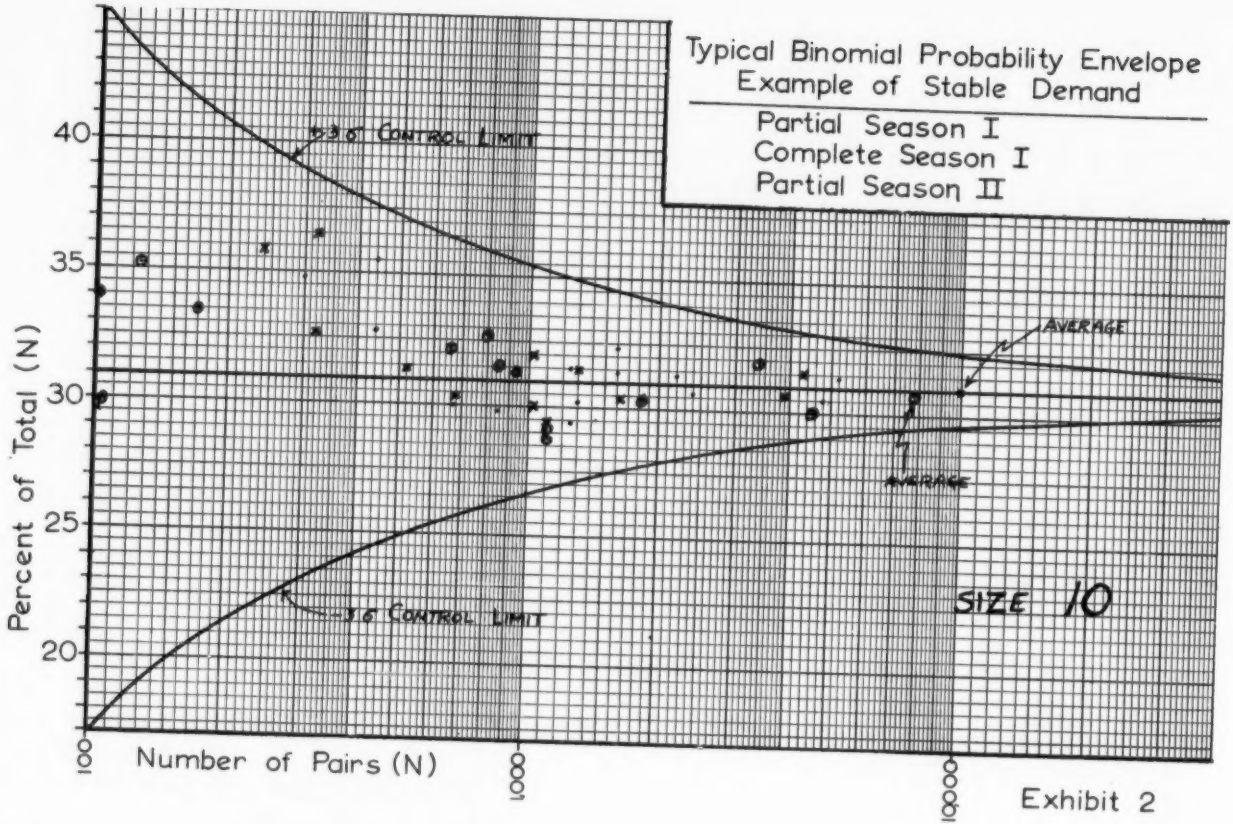

Typical Binomial Probability Envelope
Example of Stable Demand
Partial Season I
Complete Season I
Partial Season II
3 σ Control Limit
3 σ Control Limit
Average
Average
SIZE 10
Percent of Total (N)
40
35
30
25
20
100
1000
10000
Number of Pairs (N)

Exhibit 2

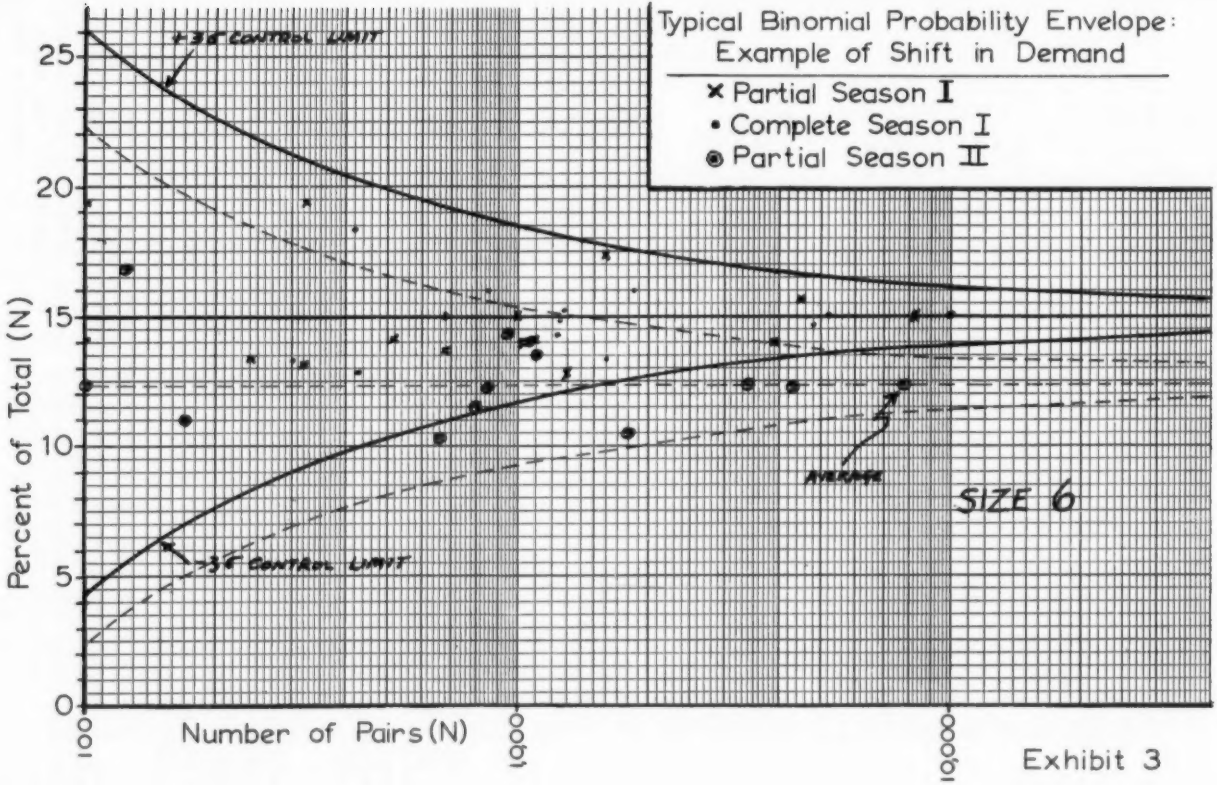

Typical Binomial Probability Envelope:
Example of Shift in Demand
× Partial Season I
• Complete Season I
◉ Partial Season II
+3σ CONTROL LIMIT
-3σ CONTROL LIMIT
AVERAGE
SIZE 6
Percent of Total (N)
25
20
15
10
5
0
100
1000
10000
Number of Pairs (N)

Exhibit 3

The meaning of this envelope is quite simple: as the total number of garments of a given lot number and color sold increases, the closer and closer should be the percentage accounted for by each size, to the true percentage. For example, if only 100 pairs are sold for a given style, fabric, and color, and the true proportion for, say, size 6, is 20%, the number of pairs actually sold in size 6 may not be exactly 20 (although this is the mean expectation), but may be as far away from 20.0% as 8.0% or 32.0%, that is, from 8 to 32 pairs, by chance alone; however, if 10,000 pairs are sold, the departure from 20.0% may be only from 18.8% to 21.2%, or from 1,880 pairs to 2,120 pairs, by chance alone. If the departure is more than these limits permit, then we must conclude that 20.0% is no longer the true proportion. Note here that in the example used, the absolue difference permitted by chance is $\pm$ 12 pairs versus $\pm$ 120 pairs, a ratio of 1 to 10, while the sample size has increased in the ratio of 1 to 100!

Exhibit 2 shows an example of such an envelope where no significant shift took place, while Exhibit 3 reveals that the proportion of sales of size 6 has moved significantly downward, to the new average and limits shown in dotted lines. This simply means that we must reduce the proportion of cuts for size 6, to compensate for this change.

When a significant shift in size preference is indicated by this evaluation, compensating cuts are computed by means of a form such as that illustrated as Exhibit 4. By using such a method, the eventual accuracy of the size distribution can be assured. Similar methods are used to compensate for color trend significance.

Pursuit of this method has produced a great many very interesting facts about the market, as well as about the opportunity for success with various styles, fabrics, and colors. The same attack has been used successfully now to permit evaluation of the significance of slower or more rapid acceptance of an item as compared with its forecasted sales, as well as drifting tastes in color preference.

In an effort to facilitate the work involved in the check point evaluations, and to make use of the charted envelopes unnecessary, we developed a set of Tables of Approximate Binomial Limits. These were run off on a computer, tabulated, and reproduced in handbook size. An example of a portion of a page of these tables is shown as Exhibit 5.

On the left hand side of each page, limits of np values are shown for both 2-sigma and 3-sigma values, as well as the value of np; on the right side of each page, with the decimal point implied immediately to the left of each number, are the corresponding percentage, or p, limits for the same confidence levels. Use of these tables makes it possible for

Exhibit 4

SIZE DISTRIBUTION EXPERIENCE REPORT

						Date	8/15/58
4	%Done	70		1	Lot No.	1298 J	
5	Std.Cst.	# 28.00		2	Color	All	
6	Price	$ 33.00		3	Forec'st	800 Doz.	

7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Size	Base per'd %	Base % x Forecast		Total Orders to date			ok or x	Distribution net change				New Base %	New % x forecast	Total cut to date	Balance to cut		OK
		Prs.	Doz.	%	Prs.	Doz.		±%	Prs.	Doz.	$				Prs.	%	
3																	
4																	
5																	
6	18.0	1,728	144 0	15.0	1,260	105 0	X	-3.0	288	24 0	671	15.0	1,440	1,080	360	10.0	ok
6x																	
8	28.0	2,688	224 0	28.5	1,960	163 1/2	OK	+0.5	43	3 1/2	100	28.0	2688	1,680	1,008	28.0	ok
10	32.0	3,072	256 0	31.5	2,240	186 8	OK	-0.5	43	3 1/2	100	32.0	3,072	1,920	1,152	32.0	ok
12	22.0	2,112	176 0	25.0	1,540	128 4	X	+3.0	288	24 0	671	25.0	2,400	1,320	1,080	30.0	ok
14																	
16																	
18																	
20																	
X	100.0	9,600	800	100.0	7,000	583 4	X	0	662	55 2	$1,542	100.0	9,600	6,000	3,600	100.0	

an unskilled office clerk to fill out such a form as that shown on Exhibit 4, at least through column number 18. Completion of the form is, and should be, a matter for executive attention.

These Tables of Approximate Binomial Limits cover values of n from 50 to 5,000, and values of p' from 0.01 to 0.50, at graduated intervals in both cases. They are useful also in quick establishment of limits for work sampling, and for other sampling applications where the conditions of the Poisson cannot be met. Copies of these tables are now available for general use.

Applications similar to that described here are being made now in a number of industries and a variety of types of problems. There is nothing difficult in use of the techniques involved. More complex developments will doubtless emerge.

The day has arrived when industries selling consumers' goods in seasonal cycles can no longer afford to take the attitude that "nothing can be done" to assist in making forecasting more scientific. What the problem needs is new and searching analysis, with the conviction that better methods are available. Many of these new methods will be adaptations of control chart methods to time series types of problems. The challenge lies in devising the means of extracting correct predicting mechanisms, and in designing economical, simple means for isolating variance that is beyond one's choice of confidence limits.

Exhibit 5

TABLES OF BINOMIAL PROBABILITY LIMITS (NORMAL CURVE APPROXIMATION) FOR 99% AND 95% CONFIDENCE LIMITS FOR p' AND np'

n Sample Size

p' True or Computed Proportion

A: AVERAGE AND LIMIT VALUES EXPRESSED AS UNITS PER SAMPLE

n	p'	LOWER LIMITS		AVERAGE	UPPER LIMITS		p'
		99%	95%	np'	95%	99%	
2000	50	930	953	1000	1047	1070	50
2000	49	910	933	980	1027	1050	49
2000	48	890	913	960	1007	1030	48
2000	47	870	893	940	987	1010	47
2000	46	850	873	920	967	990	46
2000	45	830	854	900	946	970	45
2000	44	810	834	880	926	950	44
2000	43	790	814	860	906	930	43
2000	42	771	794	840	886	909	42
2000	41	751	774	820	856	889	41
2000	40	731	754	800	846	869	40
2000	39	711	734	780	826	849	39
2000	38	692	715	760	805	828	38
2000	37	672	695	740	785	808	37
2000	36	653	675	720	765	787	36
2000	35	633	655	700	745	767	35
2000	34	613	636	680	724	747	34
2000	33	594	616	660	704	726	33
2000	32	574	596	640	684	706	32
2000	31	555	577	620	663	685	31
2000	30	536	557	600	643	664	30
2000	29	516	538	580	622	644	29
2000	28	497	518	560	602	623	28
2000	27	478	499	540	581	602	27
2000	26	458	479	520	561	582	26
2000	25	439	460	500	540	561	25
2000	24	420	440	480	520	540	24
2000	23	401	421	460	499	519	23
2000	22	382	401	440	479	498	22
2000	21	363	382	420	458	477	21
2000	20	344	363	400	437	456	20
2000	19	325	343	380	417	435	19
2000	18	306	324	360	396	414	18
2000	17	287	305	340	375	393	17
2000	16	269	286	320	354	371	16
2000	15	250	267	300	333	350	15
2000	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

B: AVERAGE AND LIMITS EXPRESSED AS PERCENTAGES OF THE SAMPLE

n	p'	LOWER LIMITS		UPPER LIMITS		p'
		99%	95%	95%	99%	
2000	50	4683	4789	5211	5317	50
2000	49	4583	4689	5111	5217	49
2000	48	4483	4589	5011	5117	48
2000	47	4383	4489	4911	5017	47
2000	46	4284	4389	4811	4916	46
2000	45	4184	4290	4710	4815	45
2000	44	4085	4190	4610	4715	44
2000	43	3986	4091	4509	4614	43
2000	42	3887	3991	4409	4513	42
2000	41	3788	3892	4308	4212	41
2000	40	3689	3793	4207	4311	40
2000	39	3591	3694	4106	4209	39
2000	38	3492	3595	4005	4108	38
2000	37	3394	3495	3904	4005	37
2000	36	3295	3397	3803	3905	36
2000	35	3197	3298	3702	3803	35
2000	34	3099	3200	3600	3701	34
2000	33	3002	3101	3499	3598	33
2000	32	2904	3003	3397	3496	32
2000	31	2807	2905	3295	3393	31
2000	30	2709	2806	3194	3291	30
2000	29	2612	2708	3092	3188	29
2000	28	2515	2610	2990	3085	28
2000	27	2418	2512	2888	2982	27
2000	26	2322	2415	2785	2878	26
2000	25	2225	2317	2683	2775	25
2000	24	2129	2220	2580	2671	24
2000	23	2033	2122	2478	2567	23
2000	22	1937	2025	2375	2463	22
2000	21	1842	1928	2272	2358	21
2000	20	1746	1831	2169	2254	20
2000	19	1651	1734	2066	2149	19
2000	18	1556	1638	1962	2044	18
2000	17	1462	1541	1859	1938	17
2000	16	1368	1445	1755	1832	16
2000	15	1274	1349	1651	1725	15
2000	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

QUALITY ENGINEERING APPLIED TO INCOMING MATERIAL

David A. Hill
Director of Reliability and Quality
Hughes Aircraft Company

Introduction

Of the group of papers sponsored by the Committee on Professional Development this one deals with quality engineering applied to incoming material. The area of interest is an important segment of the approach that A. V. Feigenbaum has called Total Quality Control (1). Incoming material means any kind of material purchased from suppliers to an organization - whether paper clips or ocean liners. Obviously the paper cannot cover all of the special circumstances that arise from such a broad range of products.

This discussion will not cover the problems of running an incoming inspection operation. The emphasis will be on quality engineering. I am going to take the stand that the whole incoming inspection operation is an excessive cost and that it is our duty to work towards abolishing it. The various approaches or programs that have been tried in recent times will be examined from this rather arbitrary viewpoint. Hopefully, some principles will be derived that can be applied by quality engineers to many situations. The ideas presented arise from the deliberations of the committee, study of published material and beliefs based on personal experience.

Some insight as to why the author is taking his position on incoming inspection may come from dropping in on such an operation. It is late afternoon on a Friday, the 27th day of September. The Quality Manager enters. He has just made his way down from the shipping area, where he has been trying to light a fire under the final test and inspection operations. Somehow half the month's production always has to be squeezed through in this last week. The big boss has had colorful things to say about meeting schedule. Now the quality manager calls the supervisor of incoming inspection out of his office and heads for the salvage crib, where suspended material is held for disposition. He picks his way among the crowded skids and the shelves packed with cartons, each dangling its tag of deceptively cheerful red. Where, he wants to know from the supervisor, are the forsaken widgets from Creep, Incorporated? Final assembly is stopped dead. There are no widgets to replace the ones that have failed in final test.

The supervisor tries to locate the widgets. No luck. An inspector is summoned to find the shipment, does so promptly and hangs around to watch the fun. The quality manager snatches the Receiving Record. The shipment was inspected to the new weighted demerit, truncated sequential, three severity level sampling plan. Not without

some difficulty, to judge by the various scratchy revisions on the Record of sample sizes and allowable defect numbers. Item: two units undersize on keyway. That means rework on the night shift after the shipment is screened 100%. Item: one unit, shaft torque over tolerance. Hopeless. That will have to go back. Are there others like it in the shipment? The torque test takes time. Also the shipment is rejected for excessive minor defects. As the quality manager stands debating whether the night shift inspection can handle screening the widgets, whether the shop still has that pick-up tool for keyways, whether to compromise on the minor defects, he sees out of the corner of his eye that Old Leatherlungs, the production superintendent, is storming down the aisle. The quality manager sighs. From the dull red of Old Leatherlungs' face he might as well prepare for the usual animated discussion of how Quality Control goes out of its way to find some fly speck faults every time a shipment is badly needed.

Let us draw a curtain on this unhappy scene and repair to some quiet place, removed from the strain and noise of production effort - say, to the quality engineer's office. The quality manager is in for it anyway. We will leave him to his problems and think about why they came about. In contrast to what we have just seen, what would be the ideal? Would it not be wonderful to get our widgets from a really competent outfit, one with a solid record of quality, whose product gives us so little trouble we can put shipments directly onto the store-room shelf? This cuts the cost of inspection for fair; never mind fancy sampling plans. Screening and rework will not hold up production. No more time spent arguing with purchasing about returning parts, faster corrective action and who pays for rework. No more debates about whether we or they measure torque correctly. An impossible ideal? Maybe. But are there some steps to be taken in that direction?

Sampling Plans Not Enough

The scene in incoming inspection suggested that statistical sampling plans for incoming inspection are not the final answer. This is not to depreciate the wide variety of sampling techniques. They do a job up to a point but they all suffer from an inescapable limit: if the required quality level is 1% defective and the material is 1.01% defective, our hopes of economizing go up in smoke. Returning shipments to the supplier does not really change this. In the long run he must recoup from his customer the cost of material returned or go out of business. In addition, when we are working in incoming inspection, we have lost two of the most powerful advantages of true quality control. First, the sequence of production is generally lost. Second, the control is too remote from the source of variations. We may be able to agree with A. F. Cowan on Acceptance Sampling, "But this is not quality control. It is the application of statistical methods to inspection". (2)

Vendor Rating

How about vendor rating? Again a wide range of plans have been adapted to various needs. (3) (4) (5) (6) They can provide for the relative importance of different kinds of defects. They compare supplier performance against a required standard. They can be simple, if that is what will do the job, or comprehensive enough to cover many hundreds of suppliers to a major program. When volume justifies, the engineer can eliminate repetitious calculations by cranking the factors into data handling equipment. Now we are getting a little closer to the benefits quality engineering can bring. The rating determines whether a supplier's product may be safely accepted after a minimum check - even skipping inspection on some lots altogether. Rating plans send up signals that corrective action is needed. An unfavorable trend over several lots is a stronger basis for action than trouble with an isolated shipment. Statistical techniques can increase the sensitivity of the index. At this point I would like to plead for the conscientious supplier. To tell him only that his deliveries for the last month were "unsatisfactory" or that some excessive number of defects was found, and stop there as some do, is simply frustrating. He needs the details of tests and measurements that provide clues to work on.

Figure 1 shows the spread of ratings for a large number of suppliers to one of our major divisions. On the charts, A, B and C vendors range from outstanding performance to acceptable; S vendors are required to demonstrate corrective action; X vendors have failed to respond to requests for action. Their orders will be cut off unless the quality manager makes a specific exception.

The quality performance index provides a guide to the selection of vendors for future orders. The quality engineer will have to realize that purchasing must also consider other factors, such as the ability to make deliveries on schedule at a favorable price. Buyers are often abused by quality people for always taking the low bid. If I were a buyer that is exactly what I would do until someone gave me sound evidence that the low bid was offset by the costs resulting from rejections, rework and customer complaints.

The vendor rating systems referenced above are based principally on the results found in incoming inspection. This is far from being all that we can learn about his performance. In fact, it can be misleading if we do not take into account what happens after the part has been accepted by the normal incoming inspection process. Occasional samples may be subjected to laboratory requalification tests that are more searching than normal incoming practice. There is the rework or rejections his material may cause during manufacturing operations. Field troubles or customer complaints are a further clue. Such data does not fit handily into statistical rating schemes but it is much too important to be left out of the total picture of vendor performance.

VENDOR QUALITY 'VQR' RATING TRENDS

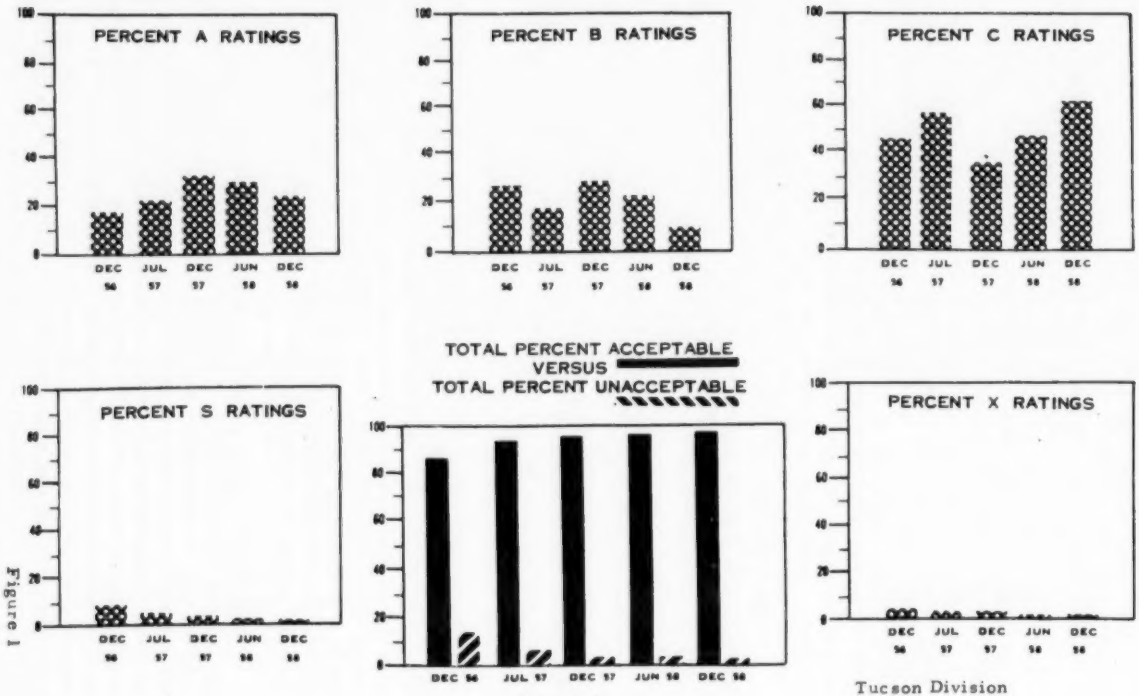


Figure 1

Tucson Division

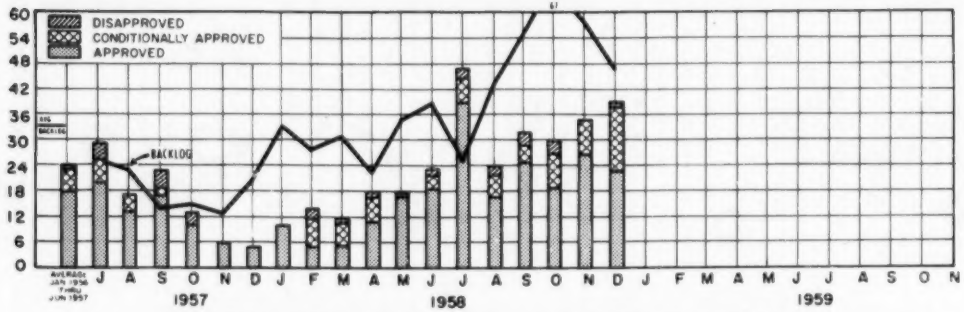
Source Selection

Rating plans, like acceptance sampling schemes, also have a limitation. They depend on accumulated experience. The fat may be sizzling on the fire before we are warned of trouble. This suggests the value of looking over new sources. Careful buyers check a supplier's ability to produce and his financial position before placing orders. There is no reason why his organization, equipment and procedures for controlling product quality should not be examined. This needs to be a lot more than taking the 75 cent tour of the plant, followed by lunch with the sales manager. Check lists can be designed for every variety of producer from buttonhole stitchers to jet engine manufacturers. They can cover very specific points. Some examples: (1) what evidence is available that test equipment is calibrated on a planned schedule? (2) does the supplier keep records of chemical and physical tests on raw materials? (3) how does he withhold discrepant material from the flow of production? There are occasions when it would be legitimate to ask for demonstration that the supplier's tooling is inherently capable of holding required tolerances. More than once our own surveys have turned up the fact that a supplier did not even possess the equipment needed to demonstrate that he was meeting some requirements of the specification. Reports need to be systematic and follow-up positive on any discrepancies noted. Figure 2 charts the survey activity for a major Hughes division. Figure 3 shows the principal issues that turned up as problems. This vendor survey procedure can increase the chance of entering into a mutually profitable relationship.

I do not want to give the impression that purchasing departments will necessarily welcome at once this added complication in making procurement decisions. We have found, however, that they will accept a survey plan once it has been demonstrated that their problems with unsatisfactory sources can be reduced. Once again we have moved a little closer to genuine quality engineering by using our heads to avoid trouble.

Again I would like to make a plea on behalf of suppliers. We quality control people are as apt as the next fellow to become arbitrary in our outlook. Our survey may reject a supplier because he does not conform to preconceived ideas on what constitutes a proper quality system. I have in mind a manufacturer of complicated and critical devices that go into advanced weapons systems. Our conscientious purchasing group was writing him stiff letters about the unsatisfactory results of quality's surveys. Since this was an important source, a visit seemed advisable. I found that the supplier maintained an extraordinarily close control over product quality. Raw materials were not just tested by batch. The variations within a strip had been found important; so the strip was chopped up and samples taken from every section. The standard practice in tooling was to cut the blueprint tolerance by a factor of ten. This insured against process variations

VENDOR SURVEY PROGRAM
SURVEY RESULTS



SUSPENDED VENDORS CURRENT STATUS (CUMULATIVE)

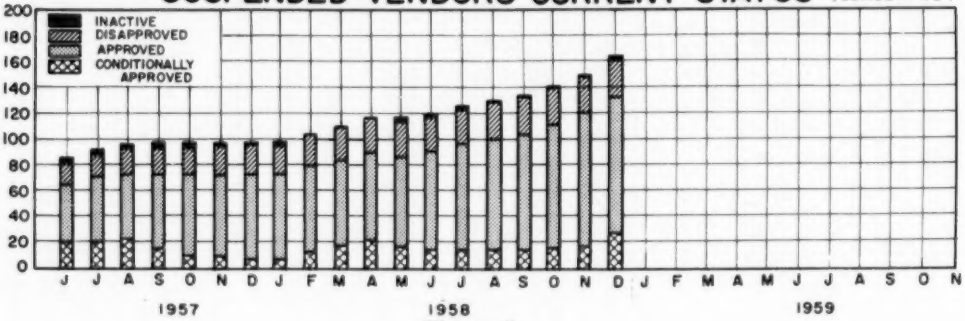


Figure 2

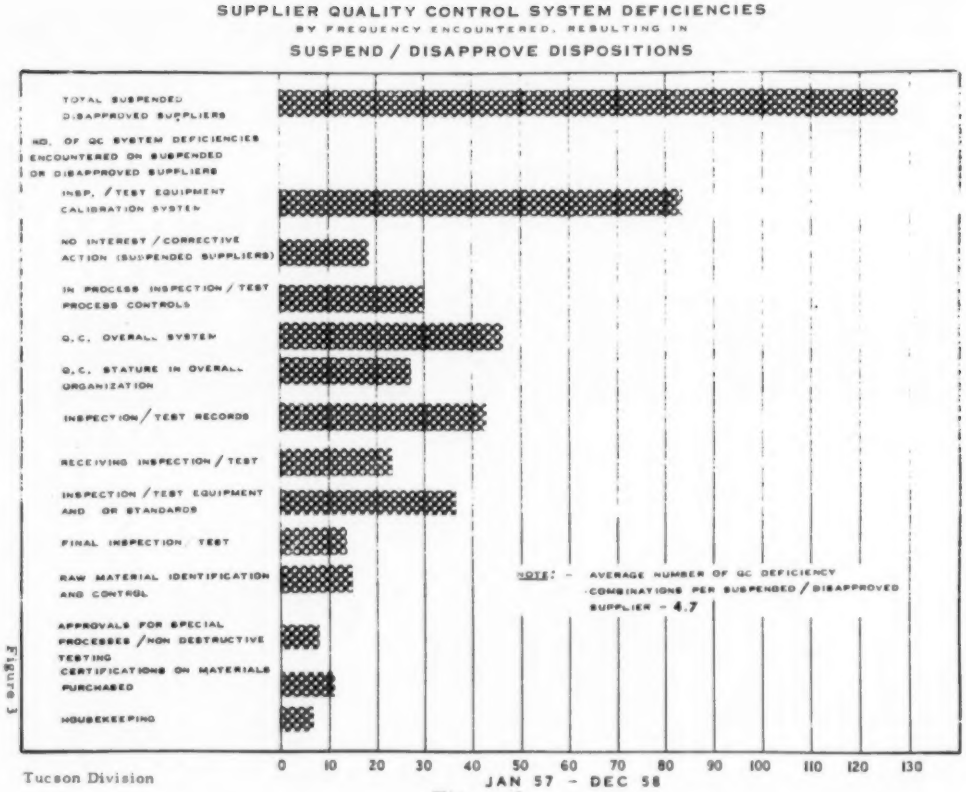

SUPPLIER QUALITY CONTROL SYSTEM DEFICIENCIES
BY FREQUENCY ENCOUNTERED, RESULTING IN
SUSPEND / DISAPPROVE DISPOSITIONS
TOTAL SUSPENDED DISAPPROVED SUPPLIERS
NO. OF QC SYSTEM DEFICIENCIES ENCOUNTERED ON SUSPENDED OR DISAPPROVED SUPPLIERS
INSP. / TEST EQUIPMENT CALIBRATION SYSTEM
NO INTEREST / CORRECTIVE ACTION (SUSPENDED SUPPLIERS)
IN PROCESS INSPECTION / TEST PROCESS CONTROLS
Q.C. OVERALL SYSTEM
Q.C. STATURE IN OVERALL ORGANIZATION
INSPECTION / TEST RECORDS
RECEIVING INSPECTION / TEST
INSPECTION / TEST EQUIPMENT AND OR STANDARDS
FINAL INSPECTION TEST
RAW MATERIAL IDENTIFICATION AND CONTROL
APPROVALS FOR SPECIAL PROCESSES / NON DESTRUCTIVE TESTING
CERTIFICATIONS ON MATERIALS PURCHASED
HOUSEKEEPING
NOTE: - AVERAGE NUMBER OF QC DEFICIENCY COMBINATIONS PER SUSPENDED / DISAPPROVED SUPPLIER - 4.7
0 10 20 30 40 50 60 70 80 90 100 110 120 130
JAN 57 - DEC 58

Figure 3

Tucson Division

Figure 3

and a build up of tolerances affecting the operation of the device. Each group of machinists was under a group leader who maintained a constant check on output with gages that were calibrated every shift. This sort of thing was true throughout the operation. It was tight in-process control justified by the nature of the product. It happened that the company was so determined to make everyone in the place feel a personal responsibility for his output that they did not have one person labelled inspector or tester or quality engineer. It was the lack of familiar labels that upset the survey team.

Requirements and Vendor Tests

From sampling through ratings to surveys we have moved back into the time cycle of procurement and closer to the source of quality. Now the question comes up, how clearly have we told the supplier what we want? Often an honest look at the cause of serious trouble with purchased material shows something lacking in the statement of requirements. An aircraft industry survey revealed that many agreed this was a problem. (7) The electronics industry has also recognized the need for clear acceptance agreements. (8)

Sometimes the problem is with blueprints and specifications. Here the quality engineer can use case histories to prove to designers what details must be included. Too often "standard practice" and "good workmanship" are cover ups for sloppy thinking. It also seems wise to tell suppliers what hurdles are planned for his product. Here I have run into occasional reluctance on the part of inspection people. One stand is that the purchase order should not say he can produce any bad parts. He must then make good on anything we find. The trouble with this is that the supplier will produce some per cent of defective material no matter how hard he tries. Most suppliers will work to an Acceptable Quality Level, formal or informal. The trouble starts when the supplier assumes that 2% is all right when the buyer wants 1%. Telling the supplier what you want seems to apply to any procurement situation. It has been systematized for a mail order house buying garments (9), in the high volume automotive industry (10) (11), and for the extremely technical devices needed by the electronics and missiles industry (12). The latter involves a fully developed Vendor Quality Certification program and a major step in the direction of our idealistic objective that incoming inspection should be abolished. In one division alone the technique has been applied to more than 30, 000 part numbers. The growth of this activity in another division is charted in Figure 4. The Q numbers refer to purchase order attachments. The supplier sends in with his shipments certified inspection and test data on a list of characteristics. He uses tests, procedures and sampling plans detailed in purchase order terms. In our business we have had to go beyond the relatively superficial "does it meet the blue-print at this moment" kind of inspection. Performance under environmental stresses and the probability of surviving for extended operation are

VENDOR QUALITY CERTIFICATION PROGRAM

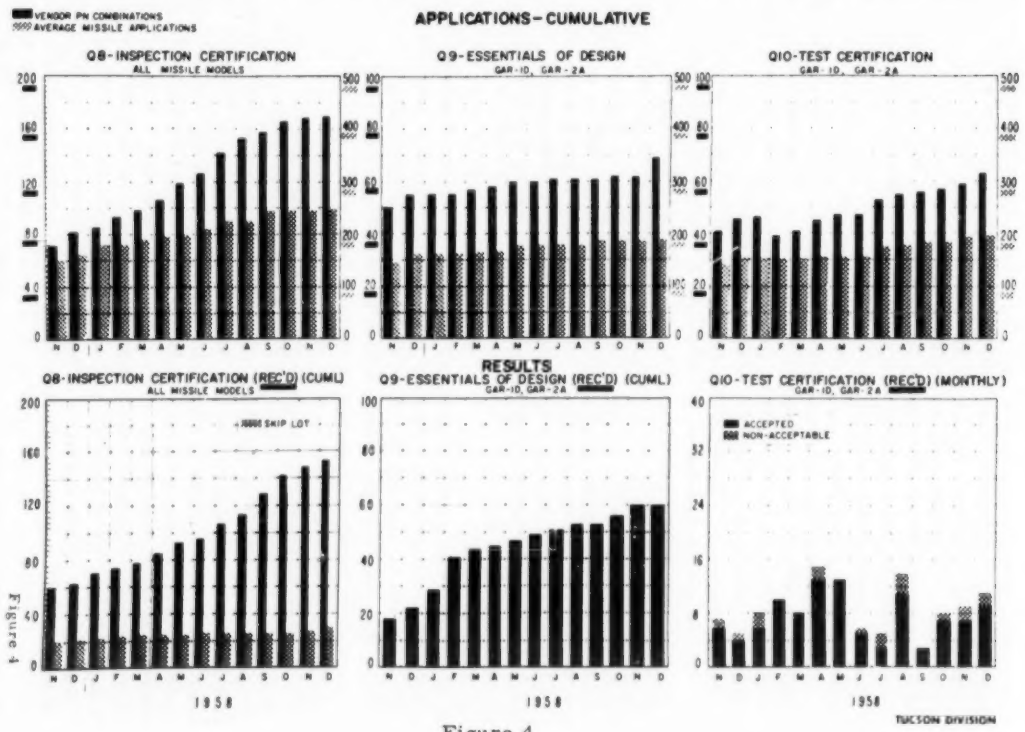


Figure 4

involved. Here the quality engineer has his work cut out for him. A set of requirements that provides ideal protection is generally impossibly expensive. He has to get the advice of the designers and use his own data to find what characteristics are the most important. A complete set of tests on a small sample for an electronic component can cost 5 to 10,000 dollars. This does not fit very well into an order for 500 parts at ten dollars each. There is plenty of room for ingenuity in designing test plans that make the best use of limited data. (13)(14) It pays to get a supplier's views on the subject. His knowledge of his product may suggest an approach that will be effective and still cost less than the requirements we develop.

Aside from the expense of tests there is the increasing problem in the defense industry of stringent failure rate requirements. A contractor for a new missile system has been inviting bids with a requirement that electronic parts have a failure rate of .001% per thousand hours of operation. Try that one on anybody's sampling tables! It is equally obvious that 100% or 200% or 300% sorting cannot guarantee such figures. The quality engineer finds himself working with his brother engineers on the safety margin between inherent parts strength and the stresses imposed by application. He will have to dig deep into the causes of process variability. We have already proved to our satisfaction that use of the Vendor Quality Certification approach can bring 10 to 1 improvements in material received and reduce incoming inspection's burden. Now we need the techniques that will bring much greater gains.

Let me hasten to say that the severe problems of weapons reliability are only an extreme case. The Vendor Quality Certification approach can be applied where no raising of quality standards is involved and the object is to maintain a satisfactory level with a minimum of wasted manpower, materials and time.

Problems of Qualification

When an order is placed for materials, the assumption is made that acceptable product received will be fully capable of meeting the user's needs. This applies whether the design belongs to the supplier or the buyer. In any industry where designs change or techniques are advancing, this often proves a dangerous assumption. The statistically minded quality engineer knows the inadequacy of most design qualification practices. It is an area where his services have not often been invited. He struggles against the commonly held impression that quality control is only some fancier form of inspection to be applied during production. There is no use wishing this bias would go away of its own accord. The invitation will come only when he has made some practical demonstrations that the designer can spare himself embarrassing troubles and that management can cut avoidable costs by including in qualification practices the techniques needed to pin down the range of

product variability. This is a cornerstone in the effort to reduce the need for incoming inspection.

Change Control and Quality

There is a second cause of trouble that also needs attention from quality people. This is change control. Changes may be made to reduce costs, to speed output or to improve performance. No matter what the reason, they tend to undermine the basis on which we may have built up confidence in a product. We see operating the well known scientific principle that if anything can go wrong it probably will. The buyer does not want to inhibit his source from making good changes or hamstring his effort with too much red tape. Sad experience has convinced more than one quality man that it is necessary to reach an agreement with the supplier that he will give notice of changes and provide evidence that it is beneficial sometimes even to the point of complete requalification. The quality engineer is not trying to undermine the authority of designers to make changes. It is his business to see that they are introduced by a planned and controlled procedure that assures continuing product performance.

Opportunities for Quality Engineering

The approaches discussed so far have been aimed at bringing material up to the receiving dock with enough confidence in its quality that it can be forwarded directly into stores. This would get Old Iron-lungs off the back of the Quality Manager. I am not a complete optimist, however. Such a millenium will not appear tomorrow, or the day after. The fast changing demands of our technical civilization will probably keep even the smartest quality engineer from being able to abolish inspection on all incoming materials. And as long as there is such an operation, quality engineers can find plenty of opportunities for their professional abilities. To suggest just two: first, we hoot at designers for setting tolerances arbitrarily. Are not most Acceptable Quality Levels set the same way? Can we find a practical method of establishing a break-even point based on dollars and cents? E. L. Grant has outlined some of the difficulties and the factors we must think about when we attempt to work on a dollars and cents basis. (15) Secondly, have we used statistical techniques to determine optimum lot sizes from a sampling, work in process inventory and overall cost viewpoint? The point is that there is more to the quality engineering world than "p" charts and operating characteristic curves. This was the approach advocated by the Professional Development Committee in the first draft of The Basic Work Elements of Quality Control Engineering. (16)

All of the methods mentioned have been tried by companies to different degrees and with varying success. Judging by our own experience it is very easy to underestimate the amount of education, persuasion and just plain patience that must go into making them effective. The

quality engineer has to reach out in two directions. Internally the engineering, purchasing and other related groups must come to see what he is doing as constructive and helpful. Without their active interest the programs will not get off the ground. The numerous pit falls involved in putting across any new Quality Control approach are a story in themselves and will not be covered here. (17)

When he has gained solid backing at home, the quality engineer must expect an equally difficult job with his vendors. Some of the ideas that appear so sound from his company's viewpoint will seem like interference and an unreasonable expense to suppliers. There is no substitute for skilled personal contact in this case. When a company has a large purchasing program and hundreds of suppliers, the contacts can be supplemented in a number of ways. My own company has used brochures, technical reports, field trips, and motion pictures to help put across the story. Groups of suppliers were given a "road show" in several cities across the country. We continue to invite suppliers to meetings held by the manufacturing divisions. They present the vendor program and specific case histories of the kinds of troubles we are trying to avoid. When a major item of equipment is to be subcontracted for design, quality control joins in the presentation made at a bidder's conference. This helps to get everyone off on an equal footing with better knowledge of what will be required.

Guide Lines for Planning

This discussion has emphasized the preventive aspects of quality control. I have borrowed an approach often stated by one of our manufacturing managers, "Someone should do something drastic". By adopting the drastic notion that incoming inspection ought to be abolished perhaps we can lift our thinking out of the rut of routine. The point of view offers a challenge to original thinking.

There are many more possibilities than this paper has been able to suggest. In selecting from the various approaches that can be applied to the area of incoming material the following principles may help:

1. We cannot expect satisfactory material from an inadequate source. Quality Control has a large stake in improving the selection process.

2. The supplier cannot be held to account for requirements that are not defined. Furthermore, specified requirements that are not backed up by a means of enforcement are so many empty words on paper. It is quality engineering's job to make sure the statement of requirements is complete, design the means of enforcement and arrive at a clear agreement with the supplier. This is a necessary service to both the engineering and procurement functions.

3. It is basic to Quality Control thinking that all controls are to be placed as close as possible to the source of variation. In the case of incoming material this means that controls maintained by the supplier will be more effective than efforts to control material after receipt. If we have done a good job of selecting a supplier, he should know more about how to control his own product than we do.

4. Purchasing has the difficult job of balancing out the factors of schedule, cost and quality for incoming material. The quality engineer can be most effective when he provides a sound quantitative basis for making the necessary managerial decisions.

An underlying purpose of these proposals is humanitarian. I would like to spare the Quality Manager from too many unhappy Friday afternoons late in the month.

BIBLIOGRAPHY

(1) Feigenbaum, A. V. "The Challenge of Total Quality Control," Industrial Quality Control, Vol. XIII, No. 11, May 1957.

(2) Cowan, A. F. "Quality Control - Is It a Management Service ?," Industrial Quality Control, Vol. XIV, No. 7, January 1958.

(3) Fitzgibbons, R. G. "The Bendix Radio Vendor Quality Rating System," Industrial Quality Control, Vol. XI, No. 8, May 1955.

(4) Georgis, G. "An Application of the Bendix Vendor Quality Rating System to a Small Plant," Industrial Quality Control, Vol. XIII, No. 9, March 1957.

(5) Wilson, S. J. "A Missiles Seriousness and Complexity - Vendor Rating System," Industrial Quality Control, Vol. XIV, No. 12, June 1958.

(6) Schneider, J. "A Vendor Quality Conformance Procedure," Industrial Quality Control, Vol. XV, No. 3, September 1958.

(7) Hill, D. A. "Vendor Quality Practices in the Aircraft Industry," Industrial Quality Control, Vol. XII, No. 1, July 1955.

(8) Quality Acceptance Bulletin Number 1, "Acceptance Agreements," Electronics Industry Association, Engineering Department, January 1956.

(9) Streit, F. "Mail Order Acceptance Sampling," Industrial Quality Control, Vol. XIII, No. 1, July 1956

(10) Smith, W. H. "Problems of Receiving Inspection and the Assembly Line," Industrial Quality Control, Vol. IX, No. 4, January 1953.

(11) Meagley, N. A. "Certification of Quality of Purchased Materials," Automotive Technical Committee, Automotive Supplement, No. 1, May 1954.

(12) Hill, D. A. and Wyllie, W. W. "Vendor Certification Program," Proceedings, Fourth National Symposium on Reliability and Quality Control, Washington, D. C., January 6-8, 1958.

(13) Davidson, W. R. "Latent Reliability - An Approximation Method," Industrial Quality Control, Vol. XIV, No. 5, November 1957.

(14) Sobel, M. and Teschendorf, J. A. "Acceptance Sampling with New Life Test Objectives," Proceedings, Fifth National Symposium on Reliability and Quality Control, Philadelphia, January 12-14, 1959.

(15) Grant, E. L. Statistical Quality Control, New York: McGraw-Hill Book Company, Inc., Chapter XVII, 1946.

(16) Feigenbaum, A. V. "The Professional Work of the Quality Control Engineer," Industrial Quality Control, Vol. XIV, No. 8, February 1958.

(17) Hill, D. A. "Communicating Quality Control Ideas," UCLA Statistics in Industry Course, 1957-1958.

"QUALITY CONTROL APPLICATIONS IN THE JOB SHOP"

W. E. Jones
Management Controls
Des Plaines, Illinois

Statistical Quality Control techniques were originated in the mass produced industries on metal parts. Their use was generally restricted to the larger companies and most small companies making many different parts in small quantities thought that these statistical techniques would not be helpful. In recent years this thinking has been changed in part. Modern Quality Control techniques can be most profitable in "job shop" production provided standard techniques are modified and applied with caution. Standard Quality Control methods must be partially set aside in order to simplify the control records. "Quick and nasty" statistical methods must be used wherever possible to save inspection and analytical time. Graphic presentation of the data is frequently set aside in favor of the data sheet. On the other hand, direct plotting of the results is frequently used where variable control is being used. In all situations, compromise is necessary to make the system easy to work with and economical to operate.

A "job shop" production run is defined as one where the time required to complete it is less than four hours or where the quantity is 100 pieces or less. The "job shop" is defined as one where more than 80% of the production runs are in this "job shop" classification. Most large plants have certain departments that fall into this classification of "job shop". A "job shop" Quality Control program is needed by almost every company, large as well as small. Most large companies have used Statistical Quality Control only on those operations that are mass produced and have avoided the "job shop" areas or departments. The large companies can make their programs more profitable by making them complete, by expanding their activities to include all "job shop" operations.

The "job shop" Quality Control program must be geared to evaluate types of operations, machine by machine. It cannot be geared to evaluate each quality characteristic on each part or product produced by each type of operation by machine. Evaluation of each type of operation rather than each quality characteristic permits the use of one fraction defective control chart which combines all the different kinds of quality characteristics and all the different part numbers. This is a "mixed up" kind of Quality Control chart; however, it will provide a Management with a record which the department foreman will find most helpful in isolating assignable causes which can be corrected to improve the overall quality of work produced. This "mixed up" kind of control chart has certain recognized short comings:

1. It does not permit defects of different degrees of importance to have more or less effect upon the chart index of quality.
2. It does not provide a sensitive statistical control which calls for corrective action as promptly as would be possible with standard control chart techniques.

3. It does not always point clearly to the source of trouble when it occurs.

On the other hand, this "mixed up" kind of machine control chart has important advantages:

1. Simplicity.
2. Low cost of records and charts.
3. It will generally, in time, show up significant differences in worn out machines or tools.
4. It will show up employees with bad work habits.
5. It will eventually show up bad design tolerances on particular parts.
6. It may even show up significant differences between Vendors; however, special supplementary studies are generally required to get this picture.

These advantages outweigh the disadvantages to justify compromise.

"Job shop" Quality Control consists of charting each machine on one continuous fraction defective control chart. Data is recorded through each 8 hour shift. Plotting is made only once each shift for each machine. The chart shows the date, employee, machine, number of inspections, number of defects observed and percentage. Part numbers should be noted only when a defect is observed. It is advisable to accumulate the results from start to end of each week so that the last plotting on Friday is actually the weekly average situation. It is not advisable to run accumulative fraction defective charting of this type beyond the one week period.

These "mixed up" machine charts should be summarized by type of machine group each week. This gives Management total performance for lathes, screw machines, headers, plating, stamping, etc. Such summary charts are generally posted near each supervisor's desk.

Percent defective is always compared with a "standard" or "allowed" number of defects computed from past performance. The foreman is requested to report on defects that exceed the "allowed" number. The foreman's report must result in corrective action or the system will be fruitless.

The Quality Control floor inspector should code each observed defect as to its cause. Following are suggested classifications:

1. Vendor.
2. Engineering design.
3. Tooling.
4. Machine.
5. Operator
6. Unknown.

This "on-the-spot" classification is used to prepare a periodic summary for administrative use. Management action should not be taken on only the evidence presented by these summaries since they represent "on-the-spot" classification. Special investigative statistical studies combined with standard average and range control charts should be applied for only limited trouble periods. Trouble periods are indicated when defects have exceeded the upper probability limits on the fraction defective machine charts or the summary charts. As soon as the combined picture returns to a satisfactory low defect

level, variables control charts, average and ranges, and/or frequency distribution analyses should be discontinued.

To staff the above system, it will require one floor inspector to each 10 to 15 machine operators. One statistical clerk will be needed for every six or seven floor inspectors. This large inspection burden may be objectionable in plants where defect cost is low and increased inspection costs are difficult to justify. In such situations, inspection, lot by lot, of finished product is recommended. Modern sampling plans should be used to provide an acceptable quality objective on all parts released to next department or to customer. Lots submitted should be kept separate by machine and by shift with sampling records noted accordingly. Summary records should be made up weekly and monthly and brought to the foreman for corrective action when defect level rises.

Miscellaneous considerations for process control follow:

1. In forming, heading and plating operations the quality variance within each part is often a clear cut probability pattern showing significant differences within each part. Stratified sampling must be used to inspect the weakest or most vulnerable area on each part.
2. Certain processes change gradually as dies and tools wear. When such trends are known to exist, there must be an exacting "first-part-inspection" and "last-part-inspection". In punch press work, the last part should be inspected and attached to the die set. Report, last part and the die must be examined by the shop foreman. Corrective work on the die must be done immediately before it is sent to storage. Such corrective work becomes a major factor in improving the total quality of work produced in a "job shop" press department.
3. A "largest and smallest" control chart is a practical way to record inspection measurements. Plotting individual measurement has its weakness. Statistical theory is not perfect for development of the upper and lower control limits on the "largest and smallest" chart. The "largest and smallest" control chart should only be used for quality characteristics that are common to all parts processed on a particular machine and where this characteristic is vitally important to the final quality. This chart will control only one quality characteristic and therefore is greatly limited as compared with the "mixed up" percent defective chart referred to earlier.
4. The "difference" control chart is another special type that is practical in a "job shop" operation. The "difference" control chart records and plots the differences between each inspection measurement and the specification mean value. Tolerances should be somewhat uniform when various parts of different shapes and sizes are being plotted from time to time. It is not advisable to change control charts every time there is a part number change. Change of part number should be recorded on the face of the chart.

5. Quality specifications for inspection must be written to standardize the inspection function itself. This is more essential on "job shop" operations due to the larger variety of parts, prints, gages, etc. On short runs it is generally too costly to make functional gages which simulate assembly use of parts.

In conclusion, modern Quality Control can be profitably applied to almost any "job shop" production in both large and small plants. It may be used as a process control program or as an acceptance sampling system. In either system, modern Quality Control will measure the attitudes of the people and evaluate the method, and provide machine and material supplied by Management. It will provide Management with a system to attain Quality Control objectives in an economical way.

SENSITIVITY AND EFFICIENCY OF EXPERIMENTAL DESIGN

A. F. Johnson
Buckeye Cellulose Corporation
Memphis, Tennessee

In much practical experimental work, the amount of effort that can be expended on a problem is limited by other than statistical considerations. The statistical problem is then to consider the most efficient manner of allocating effort and to evaluate the sensitivity of the resulting program.

The basic problem to be discussed is the estimation of a linear relationship between a dependent variable (y) and an independent variable (X).

Efficiency

The problem of efficiency is related to the selection of locations for the X variable. The efficiency of different sets of locations may be compared by the relative variances of the estimates of the slope.

Three strategies are evident from an examination of the literature. Strategy I: The X values just happened or were dug out of a file of data. This is not an experimental design, this is an accident. Strategy II: The X values are located at more or less equal increments in the X dimension in the manner of a string of pearls. Strategy III: The same number of experimental runs as in Strategy II are concentrated at a minimum number of X locations.

From geometrical considerations, Strategy III is optimum as two points determine a straight line. Psychologically, the tendency is strongly toward Strategy II as the experimenter is concerned about "missing something".

Table I gives X values for several sets of locations. These values are coded so that the range is the same in each case.

Table I
Alternate Locations for the X Variable

Number of Locations				
2	3	4	6	12
-1	-1	-1	-1	-1
				-9/11
			-3/5	-7/11
		-1/3		-5/11
			-1/5	-3/11
	0			-1/11
				+1/11
			+1/5	+3/11
		+1/3		+5/11
			+3/5	+7/11
				+9/11
+1	+1	+1	+1	+1

Now if a total of twelve experimental runs are available, any of the location patterns in Table I may be used. The variance of the slope in each case can be estimated by the following relation $\sigma_b^2 = \frac{\sigma^2}{r\Sigma X^2}$

where σ_b^2 is the variance of the slope, σ^2 the variance of an experimental run, r the number of replications and ΣX^2 the sum of squares of the X values from Table I. From this relationship, the amount of information on the slope from each arrangement is $1/\sigma_b^2 = r\Sigma X^2/\sigma^2$ and assuming σ^2 the same for all experiments the $r\Sigma X^2$ is the relative information.

Table II
Relative Information of Alternate Designs

Number of Locations	Number of Replicates	Information $r\Sigma X^2$
2	6	6.2 = 12
3	4	4.2 = 8
4	3	3.20/9 = 6.67
6	2	2.14/5 = 5.60
12	1	1.52/11 = 4.73

From the table, it is seen that the amount of information on the slope from Strategy III (2 locations) is 2.5 times as much as with Strategy II (12 locations).

More than two locations can only be justified by suspected or anticipated departures from linearity. As a general rule, an empirical polynomial expansion is considered. Under these circumstances, geometrical considerations set the minimum number of locations which thus optimizes the information. The experimenter may wish to add one additional location to permit a check for lack of fit of the polynomial of the assumed degree.

Type I Errors

The existence of an estimated slope may result from the existence of a relationship between y and X, or it may result from the random variability of the individual experimental runs. A significance test, usually by observing the relationship of the confidence limits on the slope to zero, can be performed under the usual assumptions. The $(1-\alpha)\%$ confidence limits on the slope determine the α risk of a Type I error. Since the estimate of the slope is "best" in some sense it is usually recommended that, in developing a relationship of y with X, the estimated relationship be used no matter what the outcome of the significance test. Discarding or ignoring of non-significant slopes leads to bias. In more complicated cases involving more elaborate relationships than the one under discussion, a reasonable case for eliminating non-significant terms can be made based on the practical aspects of the amount of bias introduced.

Sensitivity (Type II Errors)

In addition to the risks of Type I errors discussed in the previous section, it is possible that a relationship between y and X may be

overlooked because of an inadequate number of experimental runs. The risk of this Type II error is primarily controlled by the number of replicates of an efficient design. For a given total amount of effort, the efficiency of the design is the controlling factor. Two hypotheses are to be considered. The null hypothesis that β (the true slope) is zero and an alternate hypothesis that $\beta = \beta_1$. The estimate of β will lie within the region $\beta_0 \pm U_{\frac{\alpha}{2}} \sigma_\beta$ $(1-\alpha)\%$ of the time if the null hypothesis holds and will lie above $\beta_1 - U_{\alpha^1} \sigma_\beta$ $(1-\alpha)\%$ of the time if the alternate hypothesis is true. By combining these two regions it can be stated that if $\delta = |\beta_1 - \beta_0| \geq (U_{\frac{\alpha}{2}} + U_{\alpha^1})\, \sigma_\beta$ the risks of errors of Type I and Type II will be α and α^1 respectively. Now generally with a limited program $\sigma_\beta^2 = \frac{\sigma^2}{r \Sigma X^2}$ is given and the problem is estimating δ the minimum detectable slope using the relationship.

Table III gives the values of δ/σ for $\alpha = \alpha^1 = 5\%$ with the arrangements in Table II.

Table III

Locations	Replicates	Information	Sensitivity (δ)
2	6	12	1.04
3	4	8	1.27
4	3	6.67	1.39
6	2	5.60	1.52
12	1	4.73	1.66

The relative sensitivity is of course inversely proportional to the square root of the amount of information.

Now ordinarily the standard deviation is not known but must be estimated from the experimental data. Under these circumstances

$$\delta/\sigma = \frac{(U_{\alpha/2} + U_\alpha)}{r \Sigma X^2}$$

can be rewritten as

$$\delta/\sigma \sim \frac{t_{\alpha/2} + t_\alpha}{r \Sigma X^2}$$

This will give a conservative approximation to δ/σ. The problem in this case actually involves the non-central t distribution, but the discrepancy is only of the order of a few per cent and can thus be ignored in most practical cases where 5% risks of Type I and II errors are involved.

Efficiency and Sensitivity in More Elaborate Experiments

Considerations of the preceding sections can be applied to the design of more elaborate experiments such as polynomial expansions in several variables.

It is convenient for the less statistically sophisticated to confine themselves to experimental designs in which the various terms in the

polynomial expansion are orthogonal. Simplicity of design, computation and interpretation results from this restriction.

Table IV lists some of the characteristics of the second order composite designs, and Table V gives the sensitivity for the linear, quadratic and cross product terms of the orthogonal polynomial.

Table IV
Second Order Composite Designs

Number of Variables	Number of Terms Estimated	Number of Runs per Replicate	Comparable Replications	Residual Degrees of Freedom
2	5	9	3	21
3	9	15	2	20
4	14	25	1	10
5	20	27	1	6

Table V
Sensitivity of Second Order Composite Designs
$\alpha = \beta = 5\%$

Number of Variables	Number of Replicates	Minimum Detectable β/σ		
		β_{X_i}	$\beta_{X_i^2}$	$\beta_{X_iX_j}$
2	3	0.90	1.55	1.09
3	2	0.82	1.29	0.95
4	1	0.90	1.42	1.01
5	1	0.96	1.29	1.10

For approximately the same number of experimental runs, the second order composite designs give comparable sensitivities. It may be noted that for $\alpha = \beta = 1\%$ the minimum detectable β/σ values are about 1/3 larger than those tabulated for $\alpha = \beta = 5\%$.

The implications of the table of sensitivities are obvious for the application of the designs to "exploration" and/or "exploitation".

Conclusions

The preceding discussion has developed specifically the concepts of sensitivity and efficiency for some common experimental designs.

ON SAMPLING INSPECTION PLANS

Harold F. Dodge
Rutgers University

Consumer Protection

Early sampling inspection plans(1) were indexed in terms of LTPD (Lot Tolerance Per Cent Defective) and AOQL (Average Outgoing Quality Limit). For the first the consumer's protection is expressed as a Consumer's Risk or probability of accepting a lot of LTPD quality. For the second, the consumer's protection is expressed as an AOQL value, the upper bound to the average outgoing quality (AOQ) after inspection; specifically a per cent defective that will not be exceeded in the long run no matter what quality of product is presented, assuming screening of rejected lots and removal of all defectives therein. Of the two, the AOQL concept has found the wider use for it seems best adapted to the treatment of a flow of product or series of lots and gives to the consumer a useful index for the quality he receives. Continuous sampling plans also are generally indexed with AOQL values.

For sampling plans indexed by AQL values, such as the war-time Army Ordnance A.S.F. plans(2), the Statistical Research Group plans(3), and MIL-STD-105A(4), attention is directed primarily to the producer's interests by naming a quality level (AQL) for which the risk of rejection (the producer's risk) is small. These plans comprise a system of operations, requiring a continuing check of the submitted quality level and a switch to "tightened inspection" if the level goes bad. The consumer's protection hinges on the use of tightened inspection; failure to use the tightened inspection feature tends to render this protection uncertain and weak.

OC Curves

How a sampling plan will perform in practice is generally well indicated by its OC (operating characteristic) curve which shows the probability of acceptance for a range of values of per cent defective in presented quality. There are two basic types of OC curves which will be referred to as Type A and Type B. The type A OC curve is based on the concept of sampling from a finite universe, and shows the probability of acceptance for a lot of stated (per cent defective) quality. The Type B OC curve is based on the concept of sampling from an infinite universe, and shows the probability of acceptance for the lots coming from a process or product of stated (per cent defective) quality. The latter brings in the useful notion, employed perhaps unconsciously in control chart work, that the sample at hand is a random sample of the conceptually infinite universe of units that the process would turn out if it were to continue to operate indefinitely under the same essential conditions. There are advantages in thinking in terms of Type B OC curves, in considering the expected results of sampling from a cause system or process having a stated quality (in per cent defective).

Mathematically, Type A OC curves differ but little from Type B OC curves if the sample size (n) is not greater than 5% of the lot size (N), and for most practical purposes the two types of curves may be considered the same if $n \leq 10\%$ of N.

Sampling Plans for Destructive Tests

Sampling plans for destructive and costly tests are of necessity "small sample" plans. If sample sizes are small, it just isn't possible to have the same protection lot by lot as is obtained with the sampling plans commonly used for visual and gaging inspections. One way in which to improve the protection offered by small samples is to use the cumulative results of samples over a series of lots as a basis for acceptance, a procedure that will serve indirectly but quite effectively as a means for forcing the production of an acceptable level of quality. The use of chain sampling plans for this purpose is discussed, together with new material in preparation for publication.

References:

(1) Dodge, Harold F. and Romig, Harry G., Sampling Inspection Tables, John Wiley and Sons, Inc., New York, 2nd Edition, 1959.

(2) Army Service Forces, Ordnance Department, Industrial Service: Ordnance Inspection Handbook on Standard Inspection Procedures, Quality Control, Ord-M608-8, Issue 2, February 1944.

(3) Statistical Research Group, Columbia University: Sampling Inspection, McGraw-Hill Book Co., Inc., New York, 1948.

(4) MIL-STD-105A, Sampling Procedures and Tables for Inspection by Attributes, U.S. Govt. Printing Office, Washington, 1950.

A QUALITY CONTROL APPLICATION IN THE JOB SHOP

Robert D. Muir
Lundberg Screw Products Company

One vital segment of the metals manufacturing field is the Screw Machine Products Industry which I am privileged to represent on this panel. At the present time there are some 2000 companies in the industry with only 411, or slightly over 20%, having more than 20 employees. The average number of workers per company is about 50, so those of us who have 200 or more employed in our plant rank among the "Giants" of the industry.

My point is this: with only a few exceptions, the Chief Inspector or Quality Control Manager in a screw machine products company runs a one-man operation as far as the planning, engineering, and administering of a control program is concerned. In the face of the multitudinous duties that fall on such a member of a small company management team, how is he to help his company profit from valuable quality control plans and techniques?

Experience has shown that the "borrow, then cut and try" method is virtually a must in a job shop quality control application. One has always to be seeking the technique that is the quickest and simplest to apply and maintain, and yet will satisfy the demands of both the customer and the company watchdog on costs.

As my contribution to this panel, let me briefly outline one such application.

A large problem in the job shop - as in any manufacturing enterprise - is the consistent securing of "in print" material from a production machine. The usual procedure at the beginning of a job order is to "approve the set-up" on the basis of a "set-up sample". All this actually proves is that the set-up man has succeeded in producing a few pieces within the blueprint specifications. There is no assurance that the next index of the machine will still produce a satisfactory piece.

A brief article on standard deviation, written by Clifford Kennedy, and published in "Modern Machine Shop" almost 10 years ago, furnished the seed idea for a quick and simple solution to this problem. Borrowing his suggestion, a simple procedure and a computation form to use with it, were developed. Figures 1, 2, and 3 illustrate the application with examples from recent inspections of some machining operations in our plant.

Procedure to follow.

When a machine set-up has received "first-piece" approval and is ready to turn over to the production operator, a quality control inspector conducts a "Capability Test" on one or two of the most important characteristics as predetermined.

A consecutive sample of 20 pieces is taken (24 pieces in the case of a six-spindle screw machine). There must be no machine stoppage or adjustment during the gathering of this sample.

Each piece is measured and the actual finding is recorded on the test form. The "fineness" of measurement should be at least one-tenth of the tolerance spread. For simplicity in calculation, the measurements are reduced to just the last digits and the average ($\bar{X}$) of these readings is computed. This figure, to the nearest whole number, is subtracted from each individual measurement ($X-\bar{X}$). The resultant values are each squared $(X-\bar{X})^2$ and the average of this group of figures is computed. The square root of this average of the squares is the standard deviation of the operation being tested (s). Multiplying the standard deviation by six produces the amount of natural variation in the operation, apart from any tool wear or operator adjustment. This six standard deviations figure we recognize as the "capability" of the equipment. This figure can be compared with the tolerance specified for the operation and a quick decision made as to the ability of the set-up to hold tolerance.

Three Selected Examples.

Figure 1: The six standard deviations figure shown in the lower right corner, when compared with the .002 tolerance specified for this milled dimension, indicates immediately that there is no possibility of the print dimension being maintained. Even if the machine setting could be held at the exact mean, the natural variation of the operation will cause parts to be produced both over and under the limits. The capability (6s) of an operation should not exceed three-quarters of the tolerance if there is an honest desire to meet the specification. In the sort of situation illustrated by Figure 1 there are two possibilities open: 1) perhaps the tolerance can be opened to .003; or if not, 2) then the equipment must be improved or the operation transferred to equipment that will produce to a satisfactory capability level.

Figure 2: The capability value of this centerless grind operation slightly exceeds the recommended maximum of three-quarters tolerance, but, because the average ($\bar{X}$) is so close to the mean of the tolerance and the operator was exceptionally dependable, approval to proceed with production was given.

Figure 3: This illustrates a capability test on a six-spindle automatic screw machine so 24 measurements are shown. The capability value is within the safe limit to permit holding the .002 tolerance. However, the average of the readings ($\bar{X}$) is at 1.2466 diameter so the tool should be adjusted downward closer to the mean diameter. A most important fact must be noted here: when the computed capability of a multiple spindle operation is not within the safe operation limit of three-quarters tolerance, a new study must be made on each individual spindle separately. Perhaps only one or two of the spindles need correction, but this fact will not be discernible by the simplified test shown in Figure 3. Such stratified sampling is only permissible on a multiple operation when it proves to be capable of meeting the desired tolerance.

Any job shop, even the smallest, can greatly profit by such simplified control applications from the mysterious realm of statistics. And there are as many other applications as you desire to search out and put to work serving you.

CAPABILITY TEST

CUSTOMER — PART NO 106171

CHARACTERISTIC .749-.751 MILL TOLERANCE .002

DEPARTMENT 400 MACHINE 85 DATE 11-5-58

	(X) ACTUAL	(X) CONVERT	$(X-\bar{X})$	$(X-\bar{X})^2$	$s=\sqrt{\frac{\Sigma(X-\bar{X})^2}{n}}$
1	.7506	6	-1	1	
2	.7507	7	0	0	
3	.7509	9	2	4	
4	.7509	9	2	4	
5	.7501	1	-6	36	$s=\sqrt{13.4}$
6	.7510	10	3	9	
7	.7800	0	-7	49	
8	.7504	4	-3	9	$s=3.66$
9	.7514	14	7	49	
10	.7514	14	7	49	
11	.7501	1	-6	36	3.66
12	.7507	7	0	0	6
13	.7509	9	2	4	21.96
14	.7508	8	1	1	
15	.7509	9	2	4	
16	.7510	10	3	9	
17	.7507	7	0	0	
18	.7506	6	-1	1	
19	.7508	8	1	1	
20	.7507	7	0	0	
21					
22					
23					
24					
25					CAPABILITY (6s)
	SUM	146		268	.0022
	AVERAGE	7.3		13.4	

Figure 1 - Mill Operation

CAPABILITY TEST

CUSTOMER — PART NO 209240

CHARACTERISTIC .562-5625 DIA. TOLERANCE .0005

DEPARTMENT 600 MACHINE 55 DATE 1-5-59

	(X) ACTUAL	(X) CONVERT	$(X-\bar{X})$	$(X-\bar{X})^2$	$s=\sqrt{\frac{\Sigma(X-\bar{X})^2}{n}}$
1	.56216	16	-8	64	
2	.56220	20	-4	16	
3	.56218	18	-6	36	
4	.56234	34	10	100	$s=\sqrt{46.2}$
5	.56226	26	2	4	
6	.56228	28	4	16	
7	.56222	22	-2	4	$s=6.8$
8	.56212	12	-12	144	
9	.56218	18	-6	36	
10	.56214	14	-10	100	6.8
11	.56226	26	2	4	6
12	.56226	26	2	4	40.8
13	.56234	34	10	100	
14	.56224	24	0	0	
15	.56232	32	8	64	
16	.56230	30	6	36	
17	.56216	16	-8	64	
18	.56228	28	4	16	
19	.56220	20	-4	16	
20	.56234	34	10	100	
21					
22					
23					
24					
25					CAPABILITY (6s)
	SUM	485		924	.00041
	AVERAGE	24.2		46.2	

Figure 2 - Grind Operation

CAPABILITY TEST

CUSTOMER — PART NO 1859-B-2

CHARACTERISTIC 1.245-1.247 DIA. TOLERANCE .002

DEPARTMENT 200 MACHINE 39 DATE 9-16-58

	(X) ACTUAL	(X) CONVERT	$(X-\bar{X})$	$(X-\bar{X})^2$	$s=\sqrt{\frac{\Sigma(X-\bar{X})^2}{n}}$
1	1.2466	6	0	0	
2	1.2467	7	1	1	
3	1.2465	5	-1	1	
4	1.2468	8	2	4	$s=\sqrt{5.8}$
5	1.2468	8	2	4	
6	1.2470	10	4	16	$s=2.4$
7	1.2468	8	2	4	
8	1.2464	4	-2	4	
9	1.2469	9	3	9	2.4
10	1.2463	3	-3	9	6
11	1.2462	2	-4	16	14.4
12	1.2465	5	-1	1	
13	1.2467	7	1	1	
14	1.2462	2	-4	16	
15	1.2465	5	-1	1	
16	1.2463	3	-3	9	
17	1.2467	7	1	1	
18	1.2465	5	-1	1	
19	1.2467	7	1	1	
20	1.2466	6	0	0	
21	1.2468	8	2	4	
22	1.2463	3	-3	9	
23	1.2465	5	-1	1	
24	1.2461	1	-5	25	
25					CAPABILITY (6s)
	SUM	134		138	.0014
	AVERAGE	5.6		5.8	

Figure 3 - Screw Machine Operation

A STATISTICAL QUALITY CONTROL APPROACH TO THE SELECTION OF FLAVOR PANEL MEMBERS

Mae-Goodwin Tarver, Barbara A. Hall, and Joan G. McDonald
Continental Can Company, Inc.

Introduction:

Extraneous flavors must not alter the characteristic flavor of a food product. Physical instruments are not available at this time to detect differences in flavor and human response is still the only criterion of flavor quality. Each taster, therefore, may be considered as a measuring device used to detect differences in flavor. A sensitive taster who is able to detect small flavor differences is analogous to a vernier micrometer which can detect dimensional differences of a few thousandths of an inch. All measuring instruments must be calibrated and the precision of the measurements must be known. Since each taster is a flavor-measuring instrument, the taster's response to flavor and the precision of his judgments must be evaluated.

In appraising the reliability of a physical measuring instrument, it is customary to describe its performance by means of a capability study. The precision of the instrument is then evaluated using standard statistical methods. A recent statistical comparison of flavor scoring systems by Agnes Carlin, Oscar Kempthorne, and Joan Gordon[1] indicated that standard statistical techniques, applied to flavor scoring data, resulted in a satisfactory description of the flavor response, providing that flavor data met specific statistical requirements. Therefore, if it can be shown that flavor scoring data meet these requirements, instrument capability study methods may be used to select flavor panel members. The Two-Sample Difference Test, used in conjunction with a scoring scale, is an extension of the theory of paired comparisons. It is a statistically efficient research technique because the data measure the degree of difference between two samples and do not merely record the presence or the absence of a difference.

In the particular type of difference test, described in this paper and in a second paper now in preparation[4], two samples are presented to each taster: one is the control or standard sample, the other one is the test variate. After tasting the pair, the taster is asked to record his score for the degree of difference between the two samples on the following 0-5 subjective scale:

Difference Score	Description
0	No Difference Detected
1	Very Slight Difference
2	Slight Difference
3	Moderate Difference
4	Large Difference
5	Extremely Large Difference

Each taster is permitted to record his difference score in his characteristic numerical system. For example, one taster may record a difference score as 0.33 while a second taster may record the score for

the same pair as O.3. This individual response to flavor difference may be characteristic of a specific taster and can be reflected in his precision.

Only four samples are tasted at any one time by each taster to avoid taste fatigue[5]. Difference tests, therefore, are limited to duplicate tests of each variate to provide the required replication for measuring the precision of each taster's response. The duplicate tests consist of duplicate source samples of the control and of the test variate. Two portions from one composite control sample and two portions from one composite test variate are used in the test replication. At a specific testing period, the replicates of any one variate are randomized among the replicates containing the second variate. Hence, duplicate tests upon a particular variate are not designed to follow in sequential order.

When a Two-Sample Difference Test is run at this laboratory, the control sample of the pair is identified for the taster. This procedure is desirable as prior knowledge of the test will aid the taster to make more accurate judgments[6]. A control-to-control pair is tasted at randomly selected test periods by every member of the panel to measure any taster bias that may have been created by this prior knowledge. The tasters are aware of this procedure but are not aware of the presence of the control-to-control pair at any specific time of tasting. This check on taster bias is similar to the checks made on the accuracy of analysis in a chemical laboratory using an unsuspected standard sample.

Selection of Panel Members:

Flavor testers must be calibrated before the Two-Sample Difference Test is placed in operation. Each taster has a characteristic sensitivity level above which he will show a response to flavor and below which he will not show a response. The biological terminology for this characteristic level is tolerance[7]. The true sensitivity level associated with a flavor response may be estimated but its exact value cannot be determined experimentally. Limits can be attached to the true sensitivity level indicating the range within which the estimate is expected to fall. The word, tolerance, suggests a range or spread. Therefore, tolerance level is used in this paper to indicate the taste or flavor level associated with a taster's response.

In addition to measuring the value of tolerance levels, the reliability of each taster must be appraised since each taster may have a potentially characteristic precision. J. P. Guilford[8] has shown that the response patterns of subjective measurement data may be reflected in increased variation and frequently these are independent of the average level.

The Two-Sample Difference Test was considered for use in selecting packaging materials for products having a characteristic bitter taste, e.g., beer. Seventy-one members of the research department were screened both for their bitter tolerance levels and also for the reliability of their decisions.

The following quinine sulfate dilution series with a dilution ratio of three was used in this test for the bitter taste:

0.00 ppm quinine sulfate: Double de-ionized water - Reference point
0.10 ppm " " in " " "
0.30 ppm " " " " " "
0.90 ppm " " " " " "
2.70 ppm " " " " " "
8.10 ppm " " " " " "

Double de-ionized water was used since it was found to be more flavor-free than distilled water.

Duplicate source samples from each dilution were tested by each person participating in the test. One complete dilution series was tasted at any one time and the dilution series were always tasted in ascending order. The second sample was tasted on a separate occasion. The interval between the tests was chosen on the basis of prior experience to permit limited time effects to operate.

Each taster was asked to mark the dilution at which he first obtained a taste sensation with a plus sign (+) and to mark the preceding dilution with a minus sign (-). This method of tolerance testing is patterened after the dilution method commonly used in bacteriological work for determining the density of microorganisms[9], e.g., the MPN Method for Water Analysis.

The estimate of the bitter tolerance level was calculated as the geometric mean of the dilutions marked with a minus sign and a plus sign[8]. The geometric mean was calculated rather than the arithmetic mean because the dilution series constituted a geometric progression. The estimate of the bitter tolerance level was calculated separately for each of the duplicate tests per person.

The results of the bitter tolerance tests of the 71 potential panel members are presented in Figure 1. These data are plotted on a logarithmic scale. On this chart, the maximum sample value was plotted against the minimum sample value to show the distribution of the range of the individual tasters. The individuals who showed no variation in their test results lie on the line of zero logarithmic range.

A histogram of the logarithmic ranges of the duplicate tests is shown in Figure 2. Using a chart provided by Dr. W. J. Youden of the United States National Bureau of Standards for the distribution of ranges of duplicate samples drawn from a normal population[10], no evidence was obtained to show that the logarithmic ranges of the 71 tasters were not drawn from a normal population. This evidence is a requirement for statistical analysis. The Chi-square goodness-of-fit test of the theoretical distribution of the ranges to the distribution of the ranges obtained experimentally gave a probability of approximately 0.40.

Fourteen of the tasters screened for bitterness had been previously used on the triangular flavor test panel in this laboratory for testing experimental packaging materials. A literature search showed that research workers disagreed about the relationship between a taster's tolerance level of a basic taste and his degree of flavor discrimination in an actual food product[11][12][13]. Therefore, the percentage of times each flavor taster correctly identified the odd sample of Product A

on the Triangular Flavor Test was plotted against the taster's average tolerance to quinine sulfate. This graph is shown in Figure 3 and the tolerance level is plotted on a logarithmic scale. This chart indicates there is a relationship between the ability of these 14 tasters to identify correctly the odd sample in triangular flavor tests on beer and their tolerance levels of quinine sulfate. The precise group of seven tasters, i.e., those with zero logarithmic range, correctly identified a larger percentage of odd samples than those tasters who were less precise on the quinine sulfate test. Moreover, within each precision group, the individuals who had a low tolerance to quinine sulfate correctly identified the odd sample more frequently than those tasters who had a high tolerance.

Based on this experimental evidence, 15 tasters with an average tolerance level of 0.06 - 4.80 ppm quinine sulfate and zero logarithmic range were selected for screening for the reliability of their subjective judgment scores.

Reliability of Subjective Judgment Scores:

These 15 individuals tasted four different lots of a standard-to-standard or control-to-control comparison using the Two-Sample Difference Test for a period of five months. None of the tasters was aware of the presence of the test samples at the time of tasting. Duplicate tests were made upon each lot by each taster. A histogram of the distribution of ranges of the duplicate test scores is shown in Figure 4. The Chi-square goodness-of-fit test showed no evidence that the duplicate standard-to-standard tests were not drawn from a normal population. The probability of the Chi-square value obtained from this statistical test was again approximately 0.40. The ranges of five tasters who were shown to be mavericks on the average and range control charts were not included in the histogram shown in Figure 4. These tasters' ranges were deliberately omitted from the histogram in order to demonstrate that no evidence of nonnormality was found among the nine tasters selected as members of the flavor panel. When the maverick tasters' data were included in the distribution, the conclusions drawn from the Chi-square test were not changed. One taster was excluded from the investigation at this point because he showed no interest in flavor-testing.

The averages and ranges of the standard-to-standard difference scores for each taster were plotted on control charts as shown in Figure 5(a) and 5(b). A homogeneous and reliable taste panel for subjective difference judgments on Product A consisted of the nine tasters whose averages and ranges lay within the control chart limits. These results were confirmed by the analysis of variance given in Table I. The panel average difference when each member of the pair was the same was 0.29 and the standard error of the average of duplicate tests made by one taster was 0.2269. This meant that if all nine panel members participated in a two-sample difference test, the average difference score must be greater than 0.52 before it can be assumed that the panel is capable of distinguishing a real difference between the standard sample and the test sample. The value, 0.52, was calculated using the formula given below:

$$\bar{\bar{X}} + 3\frac{\sigma_{\bar{X}}}{\sqrt{k}} = 0.29 + 3\frac{0.2269}{\sqrt{9}} = 0.52 \qquad (1)$$

Where $\bar{\bar{X}}$ = Panel average
$\sigma_{\bar{x}}$ = Standard error of the average of duplicate tests made by one taster
k = Number of tasters

In accordance with the principle of the Null Hypothesis, a panel average difference score (based on nine panel members) smaller than 0.52 was called a chance difference of a "guess" on the part of the taster.

Panel Response to Flavor Differences:

The nine members of the taste panel were used for the Two-Sample Difference Test for a period of six months whenever Product A was tested in contact with experimental packaging materials. The experimental materials had previously passed preliminary flavor-screening tests and the Two-Sample Difference Test was used to detect small differences between the control or standard sample and the test sample. Employing the experimental design described in this paper, duplicate tests were run upon each experimental material tested.

The averages and ranges of the difference scores obtained for each taster over this six-months period were plotted on control charts as shown in Figures 6(a) and 6(b). The figures in the body of the chart indicate the frequency of occurrence of the averages and the ranges. Control chart limits from the standard-to-standard comparison were used. Ranges from some of the tasters and averages of all of the tasters lay outside of the original control chart limits. This indicated that the response of tasters to experimental packaging materials tested in contact with Product A always appeared in the average difference scores and sometimes in the range of the difference scores. The response appearing in the average was independent of the response appearing in the range, which is a measure of test variation. The independence of the average and the variation of the flavor difference scores is shown in Figures 7(a) and 7(b). For this reason, transformation of the difference to logarithms was not warranted.

The increase in range, independent of the average, when an assignable cause is present is not unique to flavor measurement. This same characteristic increase in variation is apparent when duplicate readings using a sensitive physical measuring instrument such as a vernier micrometer are compared to duplicate readings using a less sensitive instrument such as a steel rule with quarter-inch divisions. The first instrument will show characteristically less variation between the readings than will the second instrument.

Under these conditions, the theoretical F-ratio $F = \dfrac{\sigma_E^2 + \sigma_B^2}{\sigma_E^2}$

with a minimum limit if 1.00 was a satisfactory measure of the response of each taster to differences in flavor of the experimental packaging materials tested. Not only is this ratio successfully used in chemical, physical, and biological testing, but a variation of this theoretical F-ratio has been suggested for use with subjective judgment data[14]. The theoretical F-ratio measures the degree of the taster's response to the flavor differences of experimental materials and his F-ratio is based

on his particular pattern of response. Tasters who have a high response to these flavor differences will have a high F-ratio. Those tasters who have a low response to the flavor differences will have a low F-ratio.

Calculation of Theoretical F-ratios:

The theoretical F-ratio must be calculated from the components of variance. Therefore, the six-months two-sample difference test data were tested for the following requirements for applying the analysis of variance[15]:

(a) Independence of the Mean and the Standard Deviation.

The standard deviation of sets of duplicate tests was plotted against the average difference of each taster for these sets. A specimen of these graphs for individual tasters is shown in Figure 7(a). The composite graph of the results from all of the tasters is shown in Figure 7(b). There was no evidence of a significant relationship between the standard deviation and the average difference score of any taster.

(b) Normality of the Data.

A distribution of the ranges of duplicate tests was made for each taster and each distribution of the ranges obtained from the experimental data was tested against the theoretical distribution of ranges of duplicate samples drawn from a normal population. A specimen of these distributions is shown in Figure 8. No evidence was obtained from any taster's data to show that his duplicate samples were not drawn from a normal population. The summary of the Chi-square goodness-of-fit tests is given in Table II.

The fact that the flavor data obtained over the duration of this study met the criteria for the analysis of variance indicated that the nine panel members were using the degree of difference scale as though it were an infinite or continuous measurement scale. These flavor data were submitted to the analysis of variance and the components of variance were calculated from the following equations[15]:

$$MS_b = \sigma_E^2 + n\sigma_B^2 \qquad (2)$$

$$MS_e = \sigma_E^2 \qquad (3)$$

Where: MS_b = Between test material mean square
MS_e = Error mean square
σ_B^2 = Variance between test materials
σ_E^2 = Error variance
n = Number of test replications = 2

The theoretical or infinite F-ratio was calculated for each taster from the equation given below[15]:

$$F = \frac{\sigma_E^2 + \sigma_B^2}{\sigma_E^2} \qquad (4)$$

Where: F = Theoretical or infinite F-ratio

The theoretical F-ratios, based on the six-months test data are shown in Figure 9 also in Table II.

A graph was drawn, plotting the theoretical F-ratio obtained from the test data for each taster, against his average quinine sulfate tolerance in logarithms to the base 10 to show whether or not the initial panel screening for bitterness was related to flavor response to experimental packaging materials when the Two-Sample Difference Test was used. This graph is shown in Figure 10. Seven of the nine tasters showed an asymptotic relationship between the theoretical F-ratio and the average tolerance to quinine sulfate. The two tasters above and below this group were expected to be maverick tasters. This expectation was based upon prior knowledge of their flavor reactions. On the basis of Figure 10, there was evidence of a general association between taster response to experimental materials tested with Product A and the basic bitter taste measured by quinine sulfate. This same relationship was found with the triangular flavor test data.

Example of the Statistical Analysis of Test Results:

It is known that flavor data of experimental packaging materials tested in contact with beer meet the requirements for the analysis of variance when the Two-Sample Difference Test is used. Consequently, packaging materials for beer are chosen for development based on the results of the statistical analysis of the two-sample difference test data. Only those materials which are shown to be similar to the flavor-free control or standard sample are considered for potential commercial use. In the example described below, the experimental material, known as "Type X," was significantly different from the control sample at all testing periods. Therefore, this specific packaging material was rejected for use with beer.

The analysis of variance of the two-sample difference test data from "Type X" packaging material is given in Table III. A graph of the results is shown in Figure 11. The inclusion of each taster's data in the statistical analysis is controlled by a range chart for duplicate tests. This chart, shown in Figure 12, is similar to those used in the chemical industry[16] for controlling the precision of duplicate laboratory analyses.

In Table III, the sums of squares among tasters has been divided into individual degrees of freedom[15] accounting for the regression of the tasters' tolerance to bitterness as measured by quinine sulfate and for the regression of the tasters' response to experimental material flavor as measured by the theoretical F-ratio. The remaining degrees of freedom among tasters account for the unexplained variation among the tasters, that is, for the deviations from these regressions.

The mean square among tasters, in the absence of flavor response, is known from the standard-to-standard comparisons. This particular mean

square represents the random, that is, the fundamental variation among tasters. Therefore, the mean square of the unexplained variation among tasters in Table III can be divided into two portions: (a) the fundamental variation among tasters based on the standard-to-standard comparisons and (b) a remainder which contains miscellaneous factors not measured in the difference test upon a specific experimental material.

These individual degrees of freedom and the division of the unexplained mean squares are the basis for the Continental Flavor Contours established for the packaging materials tested in contact with beer. The calculation of these flavor contours is given in Table III and examples of the contours are shown in Figures 13(a), 13(b), and 13(c). The purpose of the contours is to provide an image of the flavor difference characteristics of a specific experimental material under development.

Limitations of the Taste Panel Selected by Statistical Methods:

In order to select panel members by screening for any of the four basic tastes and for the reliability of subjective judgment data using a standard-to-standard comparison, a large number of potential tasters must be available. When only a small number are available, it is statistically possible to omit screening for the basic taste. The penalty paid for this omission is that the basic taste no longer appears as a separate factor in the analysis of variance and in the flavor contours. The effect of the basic taste in this instance cannot be distinguished from the unexplained variation in the analysis and from the miscellaneous column of the contours. If the effect of the basic taste is large, its omission from the analysis as a separate factor will result in a less sensitive statistical test for the differences among storage or testing periods.

The statistical selection of a flavor panel for the Two-Sample Difference Test results in a panel of sensitive and highly reliable tasters for use in critical flavor evaluations. Such a panel of expert tasters is not a random sample drawn from the entire consumer population. This expert panel is used to detect extremely fine flavor differences and the data resulting from such an expert panel are used in research and development work. The small differences detected by the panel of experts are not expected to be noted by the majority of the consumer population. Therefore, the panel of expert tasters must not be confused with a consumer acceptance panel or used in place of such a panel.

Summary:

Statistical selection of flavor panel members for the Two-Sample Difference Test is described for experimental materials tested with a product, such as beer, having a characteristic bitter taste. The panel selection is accomplished by: (a) screening potential tasters for reliability and tolerance to the basic bitter taste using a dilution series of quinine sulfate, (b) evaluating the reliability of the subjective judgments of the panel by using a standard-to-standard comparison, and (c) using the theoretical F-ratio to evaluate the response of each panel member to the experimental materials tested. Duplicate tests made by each taster are an integral part of the experimental design and are adequate for determining whether or not the resulting data meet the requirements for the analysis of variance. Flavor contours are based upon the variation

associated with individual degrees of freedom for the tasters and upon subdividing the unexplained mean squares in the analysis of variance. The Continental Flavor Contour provides a flavor-difference image for the packaging engineer as an aid in selecting new materials for commercial use.

The panel of expert tasters selected by the statistical procedures described in this paper is not expected to be the same as a consumer acceptance panel but rather is a tool for the research and development laboratory.

Acknowledgements:

The authors wish to thank Mr. J. N. Demsey, Dr. R. E. Henry, and Dr. B. H. Morgan for their assistance during the course of this study. The authors also wish to express their appreciation to all persons participating in the flavor tests for their cooperation in this investigation.

References:

(1) Carlin, Agnes F., Kempthorne, Oscar, and Gordon, Joan. "Some Aspects of Numerical Scoring Systems in Subjective Evaluation of Foods," Food Research, 20, 3, 273 (1956)

(2) Byer, A. J. and Abrams, Dorothy. "A Comparison of the Triangular and Two-Sample Taste Test Methods," Food Technology, 7, 185 (1953)

(3) Peryam, David. "Sensory Difference Tests," Food Technology, 12, 5, 231 (1958)

(4) Hall, Barbara A., Tarver, Mae-Goodwin, and McDonald, Joan. "A New Method for Screening Flavor Panel Members and Its Application to a Two-Sample Difference Test" - In preparation for the 19th Annual Meeting of the Institute of Food Technologists, May, 1959.

(5) Laue, Elsie A., Ishler, N. H., and Bullman, Gloria. "Reliability of Taste Testing and Consumer Testing Methods. I - Fatigue in Taste Testing." Food Technology, 8, 389 (1954)

(6) Raffensberger, Elsie L. and Pilgrim, Francis J. "Knowledge of the Stimulus Variable as an Aid in Discrimination Tests," Food Technology, 10, 6, 254 (1956)

(7) Finney, D. J. Probit Analysis Haffner Publishing Company, New York (1956)

(8) Guilford, J. P. Psychometric Methods, 2nd Edition McGraw-Hill Book Co. (1954)

(9) Finney, D. J. Statistical Method in Biological Assay Haffner Publishing Company, New York (1952)

(10) Private Communication - Dr. W. J. Youden to Dr. Marjorie L. Sutherland (1954)

(11) King, F. B. "Obtaining a Panel for Judging Flavor in Foods." Food Research, 2, 207 (1937)

(12) Metzner, C. A. "Investigation of Odor and Taste. Psychological Principles." Wallerstein Laboratory Communication, 6, 5 (1943)

(13) Mackey, Andrea and Valassi, K. "Discernment of Primary Tastes in the Presence of Different Food Textures." Food Technology, 10, 238 (1956)

(14) Guilford, J. P. Fundamental Statistics in Psychology and Education, 3rd Edition McGraw-Hill Book Company, New York (1956)

(15) Snedecor, G. W. Statistical Methods, 5th Edition Iowa State College Press, Ames, Iowa (1956)

(16) Bennett, C. A. and Franklin, N. L. Statistical Methods in Chemistry and the Chemical Industry, 655, John Wiley and Sons, New York (1954)

Table I - Analysis of Variance Table for
Four Lots of Standard-To-Standard Comparisons -
Duplicate Tests on Nine Tasters

Product A

	Lot No. 1		Lot No. 2		Lot No. 3		Lot No. 4	
Source	Degrees of Freedom	Mean Square	Degrees of Freedom	Mean Square	Degrees of Freedom	Mean Square	Degrees of Freedom	Mean Square
Among Tasters	8	0.1013	7	0.1230	8	0.0816	6	0.1016
Duplicate Tests	9	0.1007	8	0.1250	9	0.1110	7	0.0714

Combined Analysis of Variance

	Degrees of Freedom	Mean Square	F-Ratio	Required F 5%	Required F 1%
Among Lots	3	0.0374	0.3699	2.93	4.54
Between Tasters Within Lots	29	0.1011 (1)	0.9806	1.79	2.33
Duplicate Tests	33	0.1031	-	-	-

Grand Average = 0.29 = $\bar{\bar{X}}_d$

Variance of Lots = $\sigma_L^2 = 0$

Variance of Testers = $\sigma_T^2 = 0$

Error Variance = $\sigma_E^2 = 0.1031$

(1) $MS_T = 0.1011$ for use in Table III

Standard Error of One Taster's Average:

$$\sigma_{\bar{X}} = \sqrt{\sigma_L^2 + \sigma_T^2 + \sigma_E^2/2} = \sqrt{0 + 0 + \frac{0.1031}{2}} = 0.2269$$

Standard Error of Grand Average of k Testers: $\frac{\sigma_{\bar{X}}}{\sqrt{k}} = \frac{0.2269}{\sqrt{k}}$

Table II - Summary of Requirements for Analysis of Variance and Theoretical F-Ratio - Six Months Test Data on Container Materials

Taster No.	No. of Sets of Duplicate Tests	Relationship Between Average and Standard Deviation Per Set	Goodness-Of-Fit-Test: Distribution of Ranges of Duplicate Tests $P[\chi^2]$	Theoretical* ∞ F-Ratio	No. of Duplicate Tests
1	12	None	0.40	3.62	24
2	8	None	0.35	1.00	17
3	13	None	0.22	2.88	27
5	6	None	0.60	2.41	13
6	10	None	0.40	1.00	20
7	10	None	0.25	3.25	20
8	13	None	0.40	8.90	26
9	11	None	0.40	2.47	22
10	10	None	0.75	1.65	20

$$^{*}_{\infty} F = \frac{\sigma_E^2 + \sigma_B^2}{\sigma_E^2}$$

Table III - Analysis of Variance Table for Data Resulting From a Two-Sample Difference Flavor Test - Type "X" Container Material Packed in Contact with Product A

	70° F Storage								
	3 Days			10 Days			30 Days		
Source	Degrees of Freedom	Mean Square	F-Ratio	Degrees of Freedom	Mean Square	F-Ratio	Degrees of Freedom	Mean Square	F-Ratio
Among Tasters	3	0.4167	-	6	0.4583	-	4	0.4844	-
Due to Regression of ∞ F	1	0.4796	4.65*	1	1.6085	15.60***	1	1.4940	14.49***
Added by Regression of Bitterness	1	0.6471	6.28*	1	0.0063	0.06	1	0.1300	1.26
Deviation from Regression or Unexplained Variation	1	0.1233	1.20	4	0.2838	2.75	2	0.1568	1.52
Duplicate Error From Standard-To-Standard (True Random Error)	33	0.1031	-	33	0.1031	-	33	0.1031	-

* = 5% Level

*** = 0.1% Level

(a) Panel Average Difference Score	1.00	1.50	0.75

Table III - Cont'd.

Components of Flavor Contours	∞ F	Bitter-ness	(1) Basic Taster	(2) Misc.	∞ F	Bitter-ness	(1) Basic Tester	Misc.	∞ F	Bitter-ness	(1) Basic Tester	(2) Misc.
(b) % of Total of Mean SQS	38.37%	51.77%	8.09%	1.77%	84.72%	0.34%	15.32%	9.62%	83.89%	7.38%	5.68%	3.05%
(c) (a) x (b) = Component of Average Score	0.38	0.52	0.08	0.02	1.27	0.01	0.08	0.14	0.63	0.06	0.04	0.02

(1) Basic Taster = MS from Variation Standard-To-Standard: Random Tasters = 0.1011 (From Table I)

(2) Calculated by Assuming MS of Standard-To-Standard Contains all Random Error of Tasters: MS Deviation - MS_T = Excess MS = Factors not Accounted for.

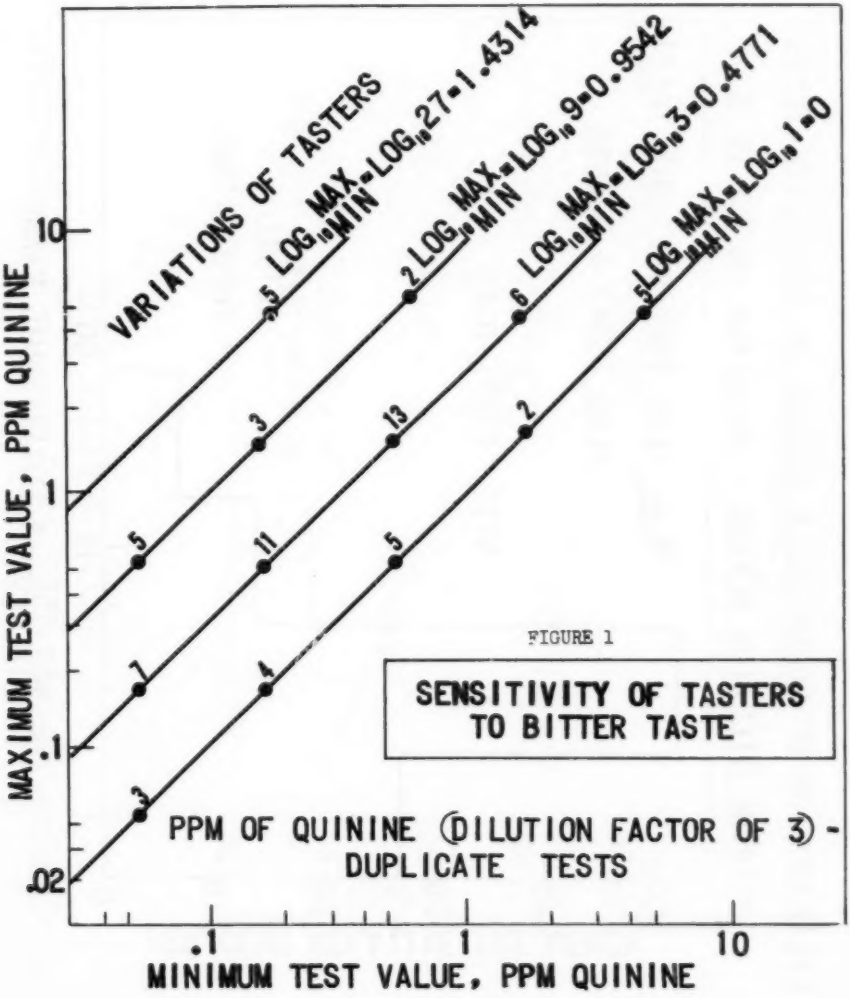


FIGURE 1

SENSITIVITY OF TASTERS TO BITTER TASTE

FIGURE 2

DISTRIBUTION OF LOGARITHMIC RANGES OF DUPLICATE TESTS

QUININE DILUTION SERIES - DILUTION FACTOR OF 3

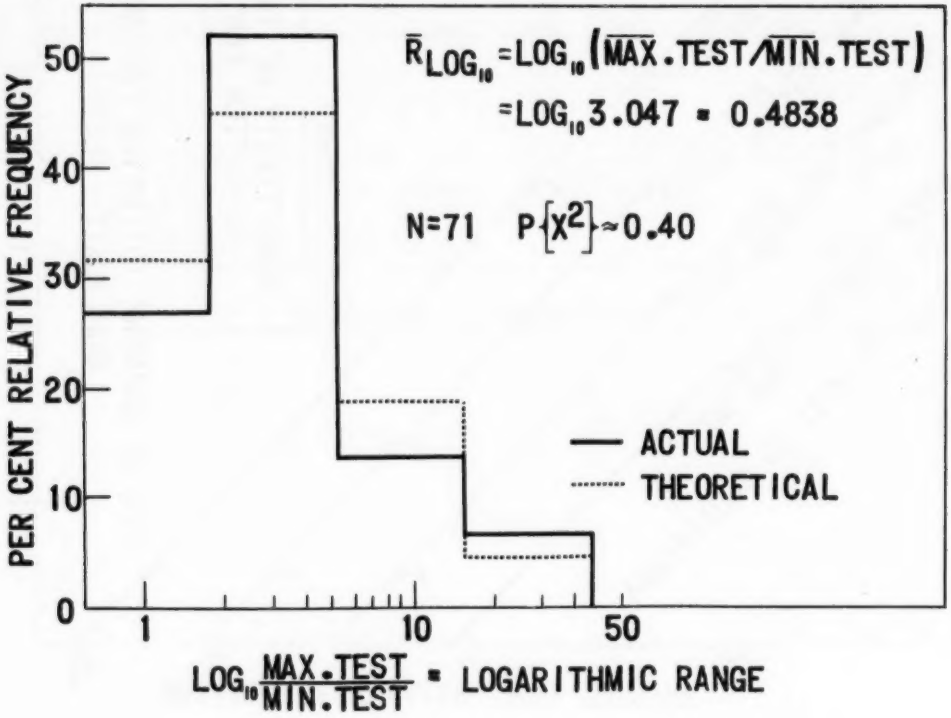


FIGURE 3

RELATIONSHIP BETWEEN TRIANGULAR FLAVOR TEST AND SENSITIVITY OF TASTERS TO BITTER TASTE

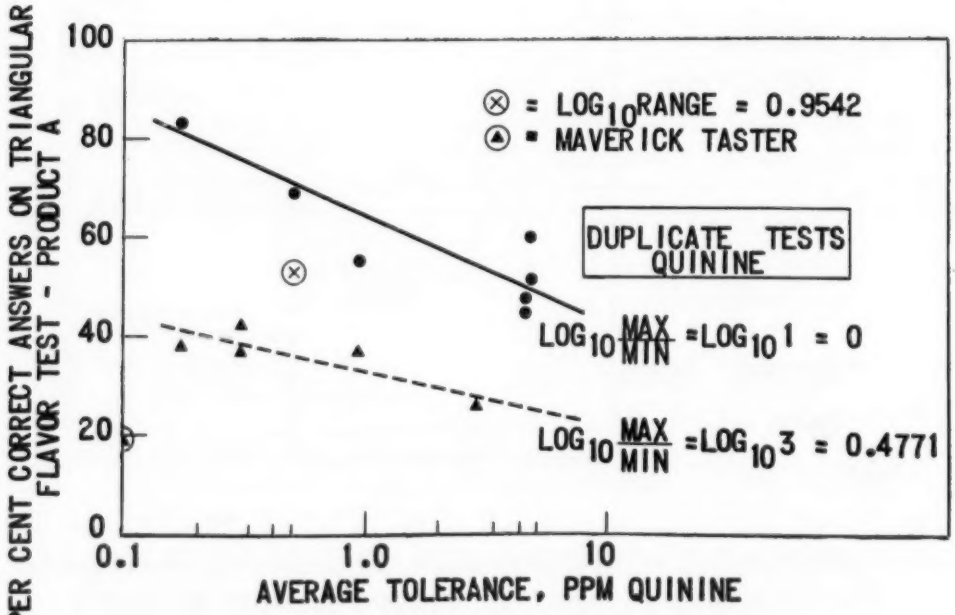
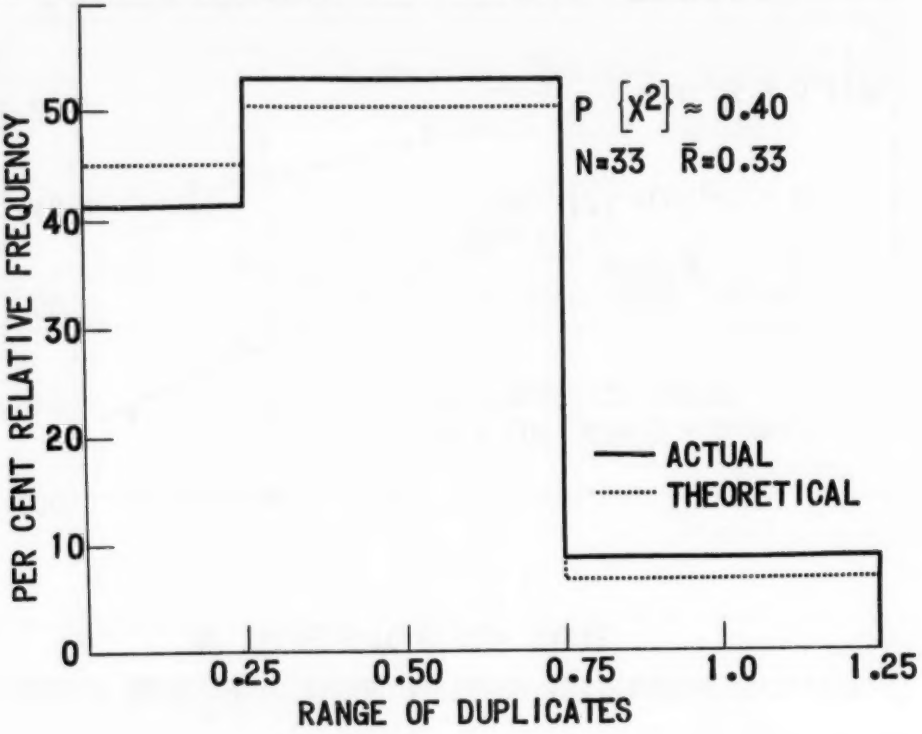
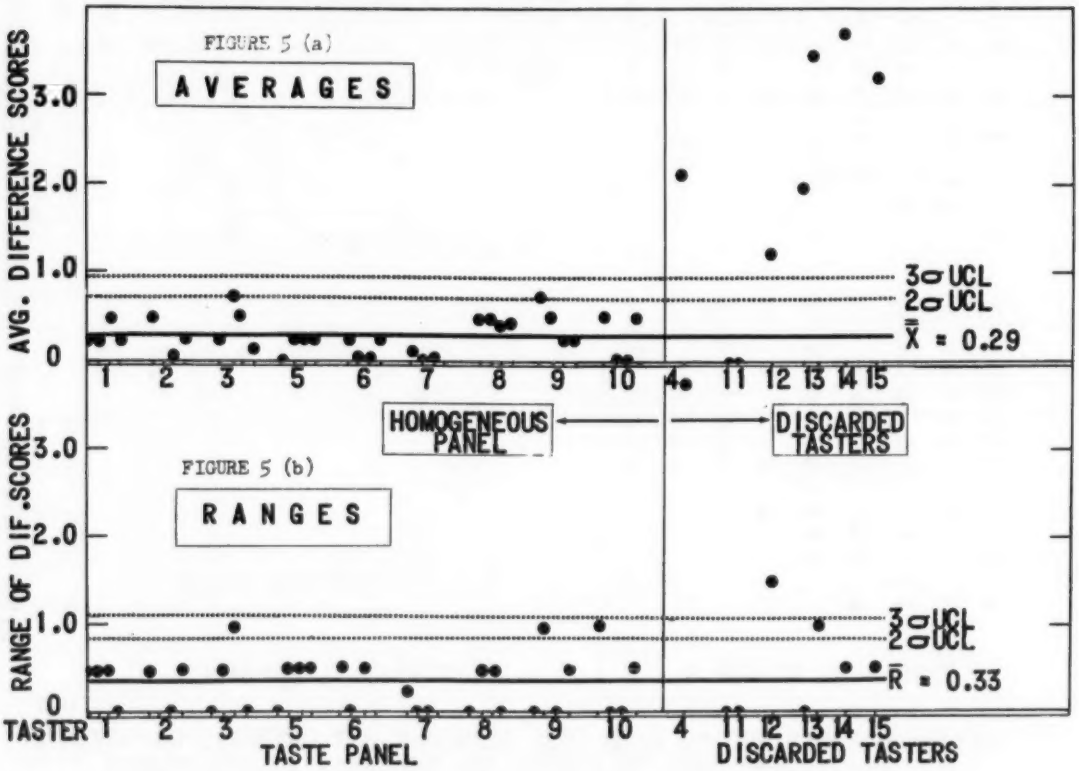


FIGURE 4

DISTRIBUTION OF RANGES OF DUPLICATE TESTS
TWO-SAMPLE DIFFERENCE TEST STANDARD-TO-STANDARD



STANDARD-TO-STANDARD - AVERAGES AND RANGES OF DUPLICATE TESTS



COMPARISON OF AVERAGES AND RANGES OF DUPLICATE TESTS STANDARD-TO-STANDARD AND 6 MONTHS TEST DATA ON CONTAINER MATERIAL

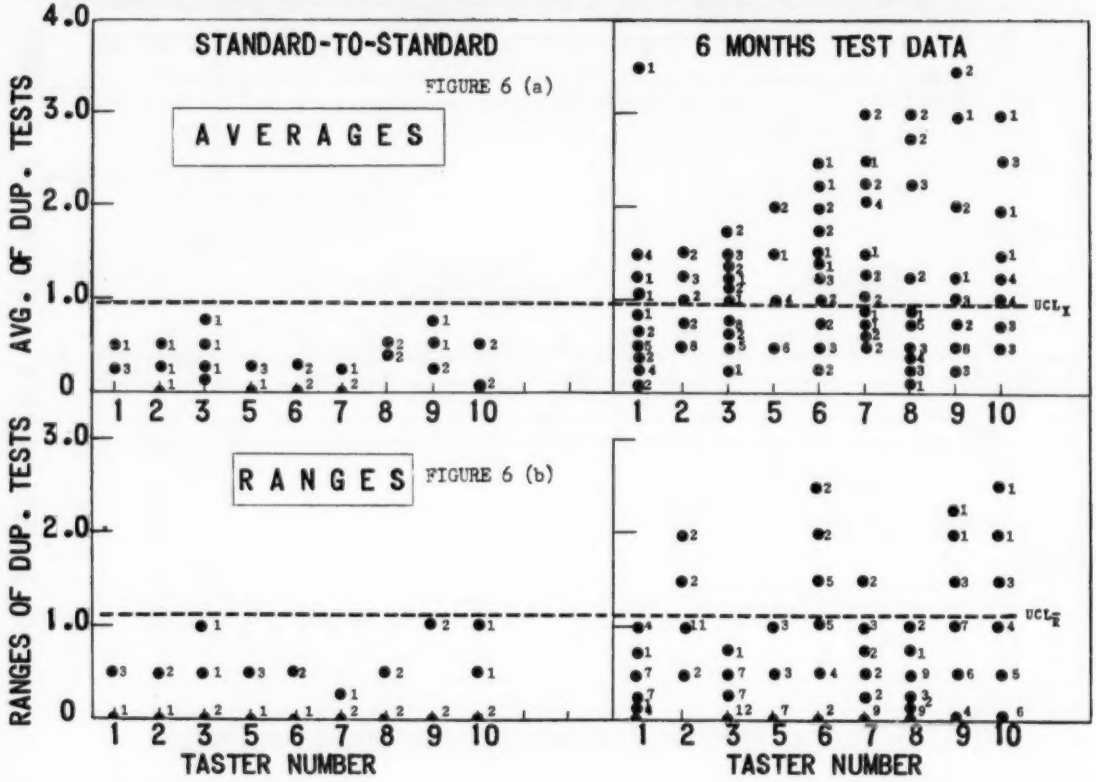
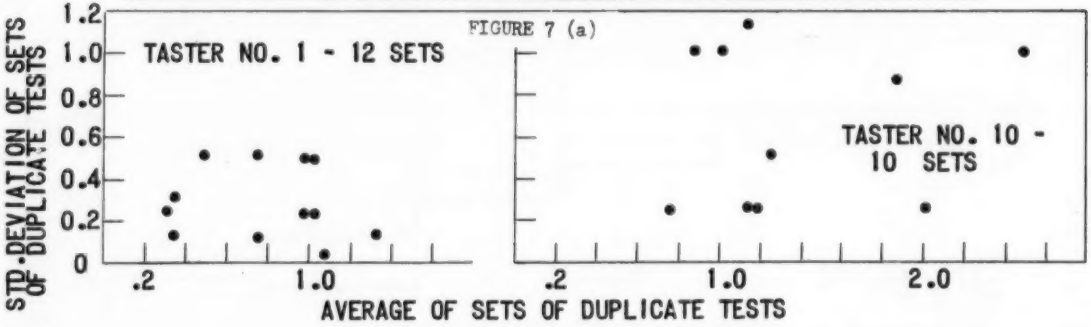


FIGURE 6 (a)

FIGURE 6 (b)

RELATIONSHIP BETWEEN STANDARD DEVIATION AND AVERAGE
SETS OF DUPLICATE TESTS - 6 MONTHS DATA ON CONTAINER MATERIALS

FIGURE 7 (a)



COMPOSITE CHART OF ALL TASTERS

FIGURE 7 (b)

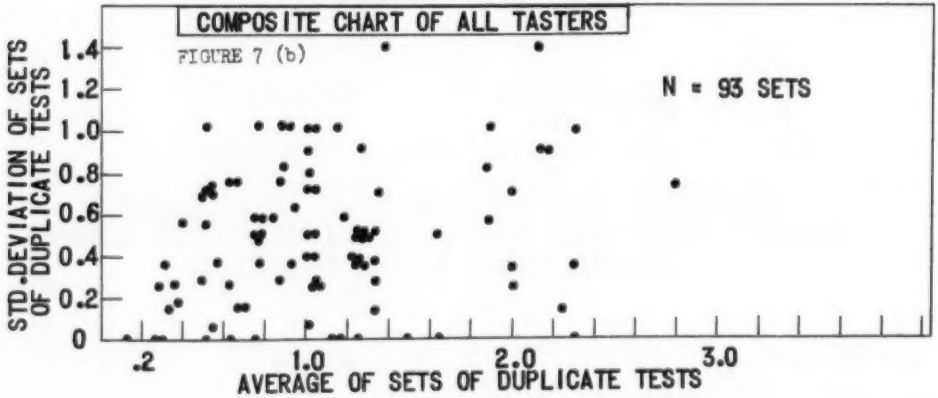


FIGURE 8

SPECIMENS OF THE DISTRIBUTION
RANGES OF DUPLICATE TESTS FOR INDIVIDUAL TASTERS
6 MONTHS TEST DATA ON CONTAINER MATERIALS

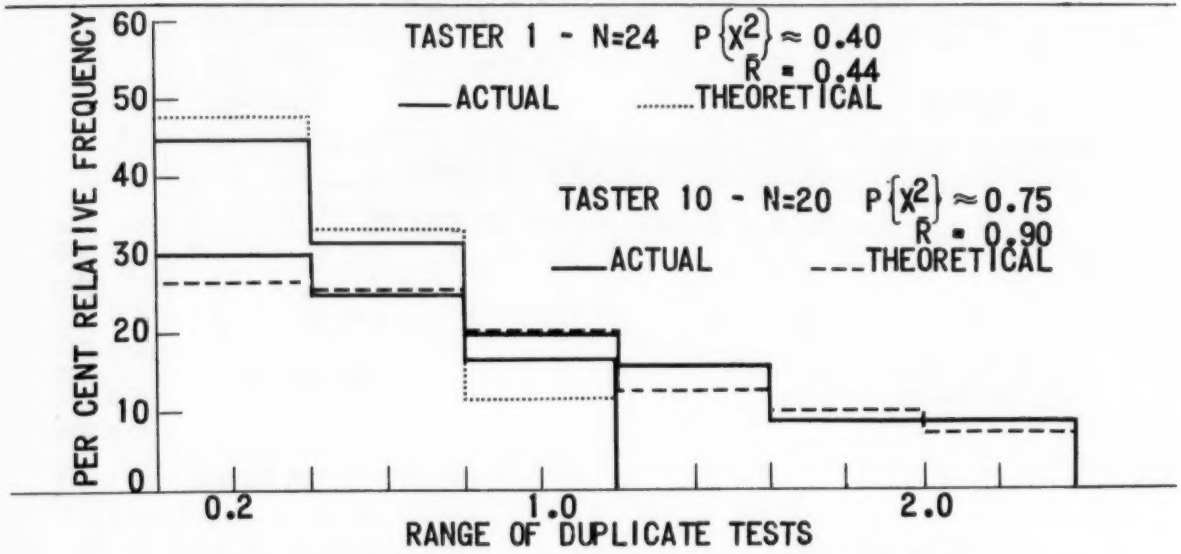


FIGURE 9

THEORETICAL "F-RATIO"
TASTER'S SENSITIVITY TO FLAVOR OF CONTAINER MATERIALS
SIX MONTHS TEST DATA

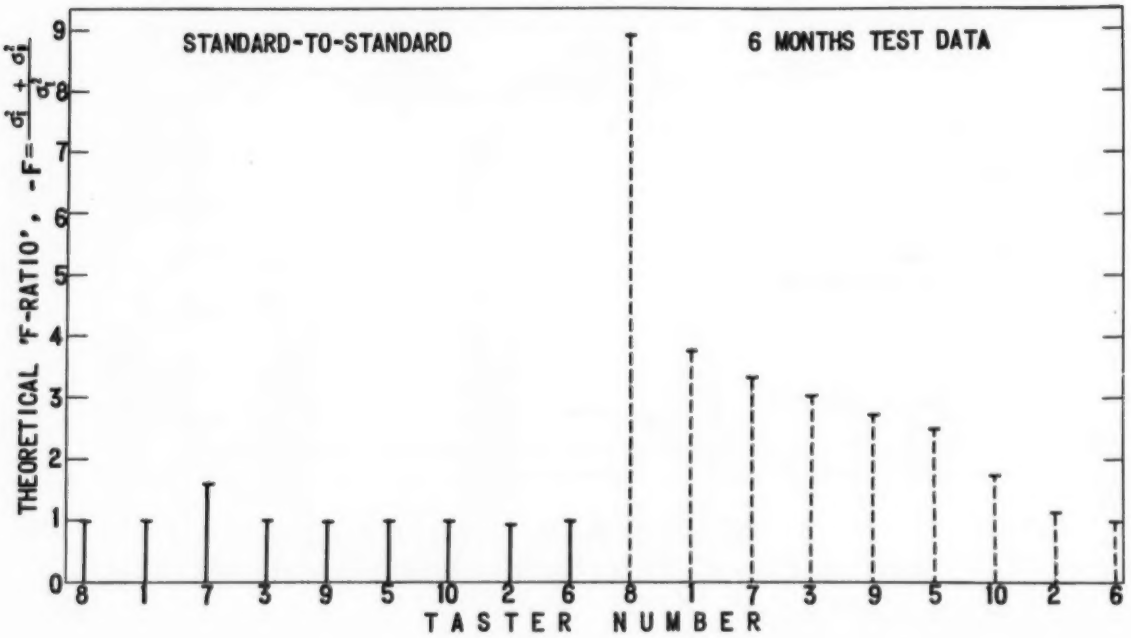


FIGURE 10

RELATIONSHIP BETWEEN THEORETICAL 'F-RATIO' AND BITTER TASTE - PPM QUININE

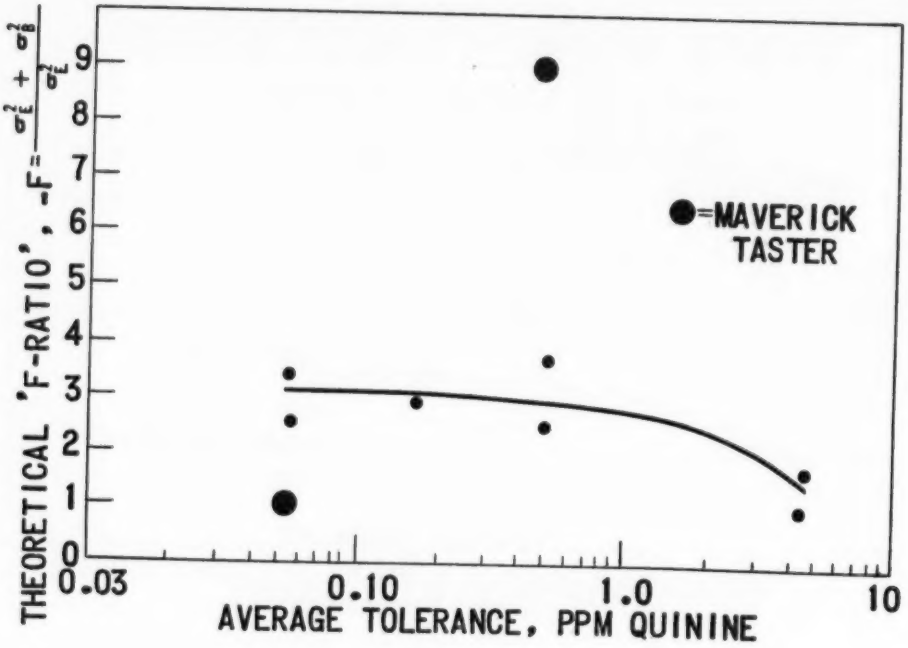


FIGURE 11

RESULTS OF A FLAVOR TEST ON AN EXPERIMENTAL CONTAINER MATERIAL

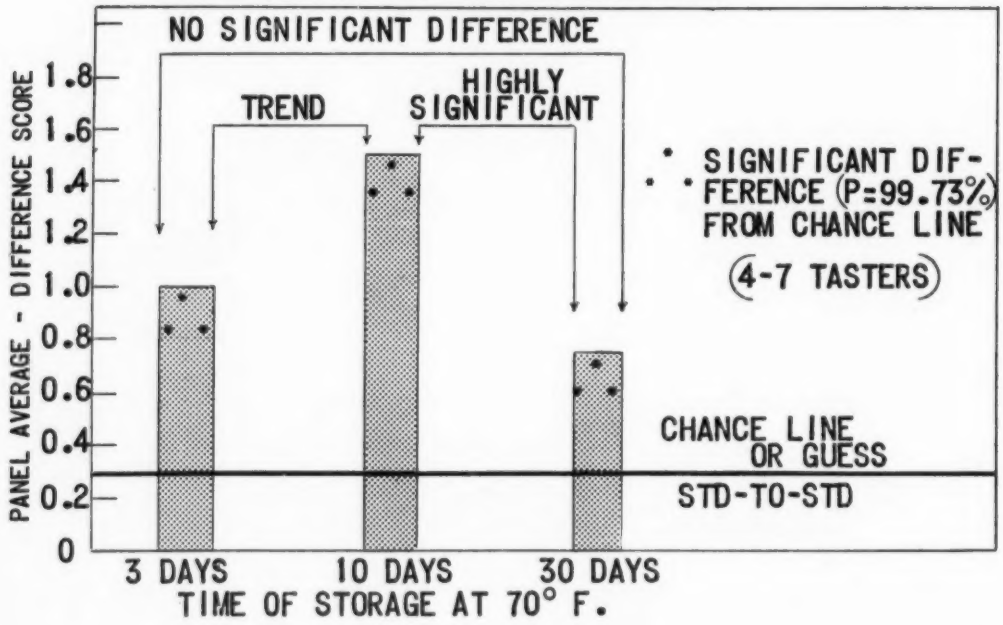


FIGURE 12 RANGE CHART TO DETECT 'WILD' FLAVOR TASTERS ON DUPLICATE TESTS

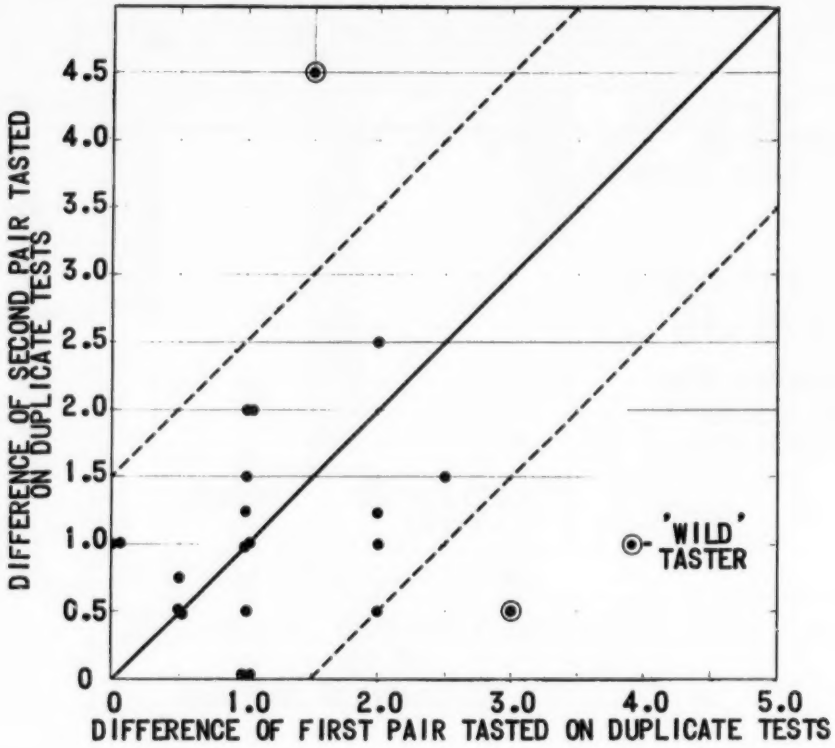
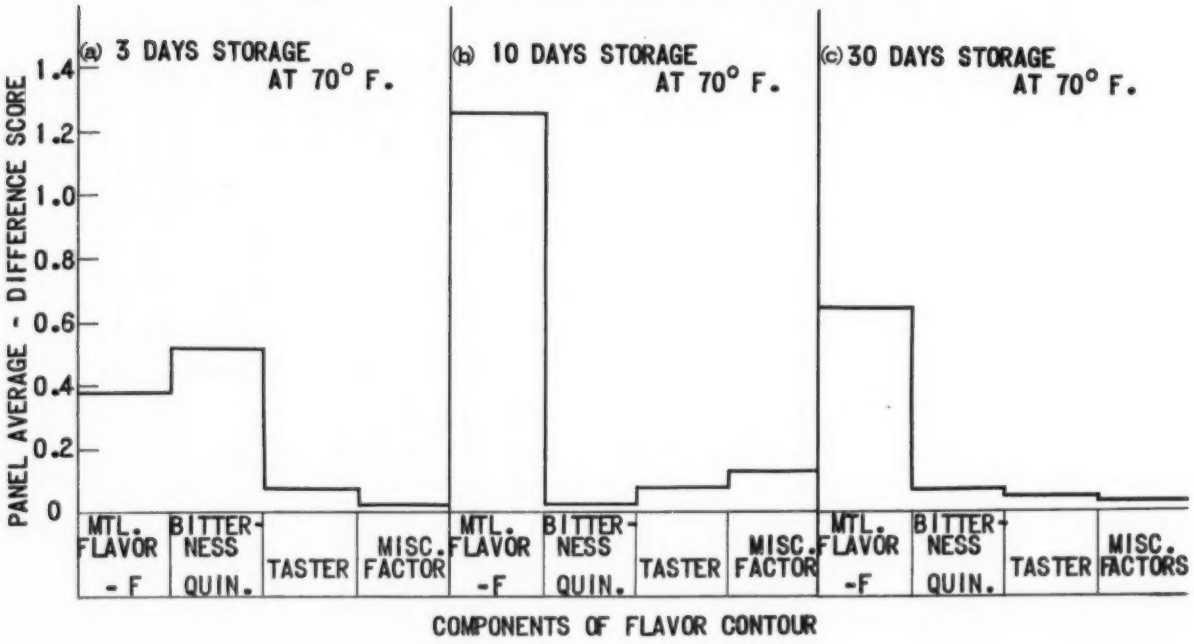


FIGURE 13

CONTINENTAL FLAVOR CONTOURS
FLAVOR TEST ON AN EXPERIMENTAL CONTAINER MATERIAL



WILL HE TAKE IT?

D. G. Meckley
The Tappan Company

To provoke your thinking and stimulate your curiosity, I have selected a problem area of inspection for which I have few answers.

When you leave, it is my intent to have you depart with a list of questions and a desire to investigate this area in your own organization.

From a selfish viewpoint, I have picked a subject which is of concern to the Tappan Company and from the panel discussion, I hope to gain understanding of the particular problem.

How many times have you heard the question, "Will he take it?"

Why is the Question asked?

The question implies many things, but it seems to me that there are two primary reasons for the question. First, it implies a lack of definition and understanding of standards; if there were completely described and understood standards, there would be no reason to ask the question. Second, it implies a leaning on the inspector and a shift of responsibility; it implies a hope that the inspector will accept a quality which is not quite up to par. Somehow we create in ourselves the idea that the inspector is endowed with some power beyond ours to separate good from bad. When he says it is good, even though we know it is not quite right, we feel released from the responsibility.

When is the Question asked?

The question is most frequently asked when the quality characteristic involved is one that must be determined by visual inspection.

In our business it most frequently involves a painted surface, a chrome plated piece or a porcelain enameled part.

For the rest of the discussion, I will call this area, the area of visual standards. It is the standard the inspector is expected to meet for the overall visual appearance of a painted piece, a chrome piece or a porcelain enameled piece, or any other characteristic of the product requiring a standard of appearance.

It is in this area where we ask our inspectors to carry one of their greatest burdens. It is in this area that we give him the least training and the most vague definition of the standard.

By Whom is the Question ask?

The question, "Will he take it?" is asked by the man on the assembly line, the foreman, the production control manager the works manager and the chief engineer. The inspector is supposed to possess some intuitive knowledge that will tell him what the consumer will accept and still stay within the practical boundaries of cost of production and company quality policies.

What is the answer to the Question?

To answer the question start by asking yourself another question, "How do we arrive at a visual standard?" Perhaps the visual standard should be set so that the general appearance is acceptable to the consumer, and since we all want to be better than our competitors, perhaps the standard should be slightly better than our competitor's standards in the same area.

To do this requires a knowledge of what the customer will accept. It requires a knowledge of the competitors' standards, and it requires an ability to define the standard. Each of these three is difficult and collectively they create the problem.

What does the Customer want?

The consumer reaction to visual standards is greatly influenced by the performance characteristics of the product; if the range performs well, the customer is usually less susceptible to complaining about small visual defects. Her feeling for the styling also accentuates or de-emphasizes the importance of visual characteristics. Certainly, the general feeling in the neighborhood for a brand or a product influences her suspicion or faith in the product and her criticism of the product. I suspect a strong regional influence in requirements for visual quality.

In many of our businesses an additional factor has come into the picture. Much of our business is now done with distributors and the actual customer reaction is once removed. When the ranges were sold only to dealers and the dealer dealt with his neighbors, our salesmen had much better touch with the actual visual requirements.

One thing is certain, the standards for the quality items involving visual decisions never stand still. They are always changing with the consumer and they are always fluctuating at the factory.

Consumer surveys have been an acceptable method of determining the customer's requirement. In the area of visual standards, I doubt if the surveys are the total answer The probability is that the persons surveyed have not owned more than one brand of the particular item and to be a real judge they should have had multiple experience with a number of brands. It is difficult to get a true reaction, when the person surveyed is not actually buying and paying for the product.

What is the correct way to determine the actual consumer requirement?

How do we know the standards of our competitors?

Here shopping the competition is a practical way to determine the quality level of the visual standards.

However, most times this shopping is done by salesmen and sometimes engineers. I doubt if one inspector in your plant has ever visited a retail sales floor. I doubt if any of them have ever talked standards to a consumer about their own or their competitors products.

How do we set the standard?

The inspector is the man who is really setting the standard in the final analysis. I said that setting the standard was the third part of

the answer. I am sure that most of us make an effort to determine the quality of the visual standard.

Probably, we include on our drawing a statement like, "Commercial quality."

Look at the other complex influences determining the inspectors concept of the visual standard.

I will guarantee that the thing he is most critical about is something involved with the current defect trend. The influence of this is fresh in his mind and this item he looks for diligently.

For a long time I have been an amateur archeologist and I have learned that when I am looking for pottery sherds, I seldom find arrow points. The reverse is true; when I look for arrow points, I seldom find pottery sherds. This is true when I am hunting the surface of a field or an area, and when I was an inspector, I noticed the same thing. The item I was particularly interested in seemed to flash out and hit me in the face, while the others would skid by. Is there a correction for this situation? Actually, we may do well to talk to psychologist to see if even the concept is correct. I believe that it is. What is your solution?

The inspector's standard is set by the opinions of the engineer and the opinion of his foremen. He hears each one of them talk about what is good and what is bad and his own standard must include some of these opinions.

Sometimes a characteristic which has never been of particular importance seems to flash into sudden prominence for some unknown reason and an acceptable existing condition that was acceptable for years becomes the focal point of the tension. How do we avoid this?

The inspector is always fearful of overriding decisions and has a tendency to be a little closer with his decisions than he actually feels the standard requires.

His previous assignment, his own work experience, and his general background affect his visual standard. Certainly, not all of us have the same taste in clothing and furniture. The inspector's taste is reflected in his standard; is this the "tase" the company wants?

His personal health and his personal like or dislike of the workers, the foremen and the engineer influences his standard.

His feeling for the supplier also tempers his visual standard.

If he has a sample, he tries to meet the sample.

The factor many times ignored is his physical ability to perceive the quality in question. Recently, we ran into the ago old problem of an electrician who could not recognize color and who was trying to wire ranges with color coded wires. A ridiculous situation, but one I suspect that is even more prevalent in the people we have looking at and determining our visual standards.

One of the most easily recognized problems is the super skill that visual inspectors soon develop. This continually forces the standard higher and higher.

I feel generally that the inspectors have a tendency to increase the level of the visual standard rather than decrease the standard. Although I am sure there are cases in the reverse.

Suggestions for better visual standards

Here are suggestions for establishing visual standards:

(1) When the standard is to be established, hold a panel discussion among:

(a) Engineering.
(b) Sales.
(c) Service.
(d) Inspection.

Try to generally describe the standard.

(2) Be sure the inspector understands that he should look at the part as it will appear in the range. This is a simple concept, but is probably the most important item in the visual standards.

(3) Be sure that the inspector knows that the visual standards are to be set to meet the consumer's acceptance. His idea of what the standard should be is not important.

(4) Use the scientific laboratory tools to at least define those items which can be defined so that the inspector has the knowledge that in a case of question he can refer to the laboratory. This will have some tendency to prevent him from becoming more and more stringent.

(5) Try to retain samples where this is feasible. The sample route often is prohibitive. In the case of chrome parts, this would require a dozen samples to describe most of the possible situations, and even then many would be missed.

(6) For very important items, send a salesman-inspector engineer team into the field to attempt to "get the feel" of consumer demand.

(7) Use available information from testing laboratories and magazine laboratories. These people see a good cross-section of products.

Suggestions for maintaining standards

(1) Periodically have the group of inspectors inspect parts together and discuss the acceptability. This should help to maintain consistency and to bring the standard requirement back to focus.

(2) Occasionally, get panels of wives and local consumers to vote on the acceptability or non-acceptability of visual items.

(3) Send the inspectors on trips to the retail floors and service organizations.

(4) Rotate the inspectors so that super skill is not developed.

The big question really is how do we get the inspector to think like a customer thinks and keep this thinking at one plateau.

The balance of quality effort

Today there are very narrow perimeters for the manufacturing cost of appliances only so much can be spent to make the product and the question is within that perimeter, "Where do we stress the quality?"

How much stress should the visual quality receive as opposed to the functional quality?

Stoves have progressed from cast iron ranges to highly styled ranges involving a rather complex electrical system. How do we convert inspectors from visual inspectors en toto to inspectors who are equally interested in the functional quality?

I would hazard a guess that in most of our plants we are far more critical than the consumer.

There is no intent to suggest sending the American public a lesser quality item. Rather the intent is to suggest that we balance the effort between functional and visual quality and that we recognize the customer requirement as the controlling factor in visual standards.

Of all the standards we set, the visual standards are the slipperiest. We have no real solution to their proper establishment.

We are very much aware of the quality being like your personal reputation. It takes years to build and hours to destroy. Sometimes its destruction comes from the smallest and unfairly stated remark. The same is true with all our products. Getting careless with visual standards can do the same thing.

What is my conclusion? My conclusion is that visual standards are tough to set and tougher to maintain.

IMPLEMENTATION OF NEW CONCEPTS IN THE INSPECTION OF SUBSISTENCE

Elwin R. Prather
Staff Veterinarian
Office of The Quartermaster General
Department of the Army
Washington, D.C.

I would like first to briefly describe the procedures employed to inspect subsistence prior to 1957, when the policy of requiring the contractor to perform this inspection, was initiated.

During World War II the inspection of subsistence was accomplished by inspectors being present in the contractor's plant to observe all phases of production, examine all components and end items on an "in-process" basis and arrange for laboratory testing in a Government Laboratory. This method of inspection required large numbers of Government Inspectors, was costly and many times duplicated inspection systems employed by the contractor. Philosophy adopted after World War II was to provide a product specification which contained minimum requirements, as near commercial production as possible, with emphasis to describe the end item. It contained a section entitled "Quality Assurance Provisions" which provided complete procedures for inspection including a Defects Classification, sampling plans and Acceptable Quality Levels. Concentrated efforts to develop these type specifications for subsistence items has been underway by the Quartermaster Food and Container Institute and the Military Subsistence Supply Agency for several years.

By 1957 inspection by Government Inspectors of the end item by use of definitive "Quality Assurance Provisions" in the specification or contract had been extended to most ration items with military characteristics, which were procured by the use of Military Specifications. However, many items procured by the Military Subsistence Supply Agency are commercial type items procured by use of a Federal Specification. Included are many perishable meat items inspected by the Army Veterinary Corps and the Meat Inspection Division, United States Department of Agriculture by the complete "in-process" type of inspection to determine contract compliance. Other items are carcass meat, fresh and canned fruits and vegetables, grain products, poultry and eggs and some dairy items. The requirements in these Federal Specifications consist largely of a Federal Grade or Standard. Inspection of these items for the Military Subsistence Supply Agency is accomplished primarily by the various Grading Branches of the Agricultural Marketing Service, United States Department of Agriculture.

For the Military Subsistence Supply Agency to implement the policy requiring the contractor to perform the inspection and testing of subsistence it was obvious that all specifications must contain complete and definitive "Quality Assurance Provisions" which outline exactly to the contractor the examinations and testing required, methods of recording results, etc.

As all specifications, both Military and Federal, contain the exact laboratory testing required and methods of testing it was decided to divide this program into two phases.

PHASE I - Contractor's Laboratory Testing Program.
PHASE II - Contractor's Inspection Program.

Phase I - the Contractor's Laboratory Testing Program was implemented late in 1957 and early 1958. Contract clauses were developed which provided the specific tests to be accomplished, methods of conducting these tests, number of sample units to be tested and method of recording results of tests. A manual entitled "Military Subsistence Supply Agency Testing Procedures Manual", which explained the contractor's responsibilities in this program was also developed and distributed to potential contractors. As industry's availability of adequate testing facilities for all subsistence items was not known the early contracts for many items were written for voluntary participation. Response to this program was good. Government Inspectors will accept the product for the testing requirements specified based on a satisfactory contractor labotatory test result. The inspector selects samples for Government verification testing based on frequency criteria furnished by the Military Subsistence Supply Agency. Results of contractors testing and Government verification testing are submitted to the Military Subsistence Supply Agency Subsistence Testing Laboratory in Chicago which statistically determines reliability of each contractors laboratory testing. Approximately three hundred contractors are now participating in this program. Reliability of these contractor's testing has been extremely good which has allowed the frequency of Government verification testing to be reduced for many items. It is anticipated that mandatory contractor testing will be extended to nearly all subsistence items during the calendar year 1959.

Implementation of the Second Phase of this program requiring the contractor to perform the actual gross inspection of the products concerned has proceeded slower than the laboratory testing program. As previously mentioned the contract must contain complete detailed inspection instructions for the contractor to follow so as to enable him to perform these inspections. The Military Subsistence Supply Agency has developed a manual entitled "Contractor Inspection Procedures Manual" for use by the contractors participating in this program. It contains general information regarding the program and defines inspection terms customarily used. It explains how the sampling procedures and Tables of Inspection by Attributes in Military-Standard-105 are applied. It provides the manner in which the results of these inspections will be reported. This manual will be referenced in contracts.

The detailed instructions for inspection to be performed are contained in the contract specifications Section 4 entitled "Quality Assurance Provisions". These provisions provide -

1. Exactly what will be examined; the end item, component or both.
2. Sample unit to be examined - pound of product, can of product, individual piece of product, etc.
3. The sample size - the number of sample units to be examined. It is designated by a term "Inspection Level" which when applied to the Sampling Tables of Military-Standard-105 gives the sample size for any sized lot.
4. Classification of Defects - this is a listing of all defects classified according to severity. Critical, major and minor is a common classification used.
5. An "Acceptable Quality Level" is specified for each class of

defects. This is the number of defects that may be present and the product acceptable for the purpose being procured.

6. Requirements for laboratory testing are included which heretofore were contained in special clauses of the contract.

7. May provide certain inspection to be accomplished by the Government Inspector during processing or between processing steps. These inspections would be for factors that cannot be verified by the Government Inspector in the end item.

Here is an example of how this will work -

X and Company has been awarded a contract for a nonperishable subsistence item to be packed in a certain size can. A lot of 7000 cans has been produced. For an item of this type four examinations are usually specified. Examination (1) - Primary container (filled can); Examination (2) - Net weight; Examination (3) - Packaging characteristics (vacuum and interior of the can); Examination (4) - Product characteristics (color, consistency, foreign material, etc.).

The sample unit to be examined is specified as 1-unit container (1 can). To determine the sample size Inspection Level I, Military-Standard-105 is specified for Examination I and Inspection Level L-5 Appendix to Military-Standard-105 for Examinations II, III and IV. When the inspection levels are applied to Military-Standard-105 we find that 150 cans are to be selected for examination in Examination I and 25 cans for Examinations II, III, IV. These cans are selected at random from the lot and carefully examined. Defects as listed for each examination are to be recorded. An Acceptable Quality Level is listed for each examination.

Let us suppose that for Examination I and Acceptable Quality Level of 1.5 defects per 100 units is specified for major defects and an Acceptable Quality Level of 4.0 defects per 100 units for total defects. Converted into acceptance numbers from Table IV-A Military-Standard-105 a sample lot of 150 cans can have 5 major defects and 11 total defects present and the product be acceptable. Any more than this number of defects found would indicate an unsatisfactory product from the standpoint of Examination I, the primary container. This same procedure is applied to the other examinations. When all examinations are completed, defects allowed by the Acceptable Quality Level to determine whether the X and Company's Quality Control System has been adequate to produce an acceptable product. If laboratory testing is also required the contractor would select the sample units to be tested in their own laboratory or a commercial laboratory retained for that purpose. The results of these examinations and laboratory testing of this and other lots making up a contract or shipment would be given to the Government Inspector at a time and place previously agreed upon or specified in the contract.

The Government Inspector will verify the contractor's inspection results by one of several methods. If the inspector is already in the plant for another purpose such as other intermediate inspection of components or factors he is unable to verify in the end item, he might conduct his verification inspection by observing the contractor's examination of the end product. If he arrives at the plant after contractor's inspection is completed, and the product is in the warehouse, he will select samples and conduct the various examinations by using the same

crieteria as specified for the contractor.

As the Government Inspector has a responsibility to accept only products that fully comp'ies with the contract requirements the inspector must do verification inspection at a frequency rate to assure acceptance of compliance product. The extent to which the frequency rate of verification inspection can be reduced is entirely up to the food industry. Inspection will be reduced as the Contractor's Quality Control Systems become fully effective. As the Contractor's Quality History improves verification inspection may go from Type A - verification on each lot, to Type B - reduced number of samples per lot or verification on a skip-lot basis to the ultimate Type C which is primarily the review of the contractor's inspection records with less frequent verification inspection and testing.

QUALITY PROBLEMS ENCOUNTERED IN THE NAVAL NUCLEAR PROPULSION PROGRAM

Donald J. Greb
Assistant Quality Control Manager, Bettis Plant
Westinghouse Electric Corporation

The quality problem in the Naval Nuclear Propulsion Program has been that of getting the various parts of the power plant to the shipyards in acceptable condition. Unsatisfactory or doubtful quality manifests itself in several ways:

1. Return of equipment to suppliers.

2. Major repair programs carried out at the shipyard, usually paid for and physically carried out by the original equipment supplier.

3. Excessive malfunctions during the testing period prior to acceptance of the power plant by the Navy.

4. The necessity for performing additional tests and inspection on equipment already delivered or even installed, e.g., x-ray of valves already welded into the system.

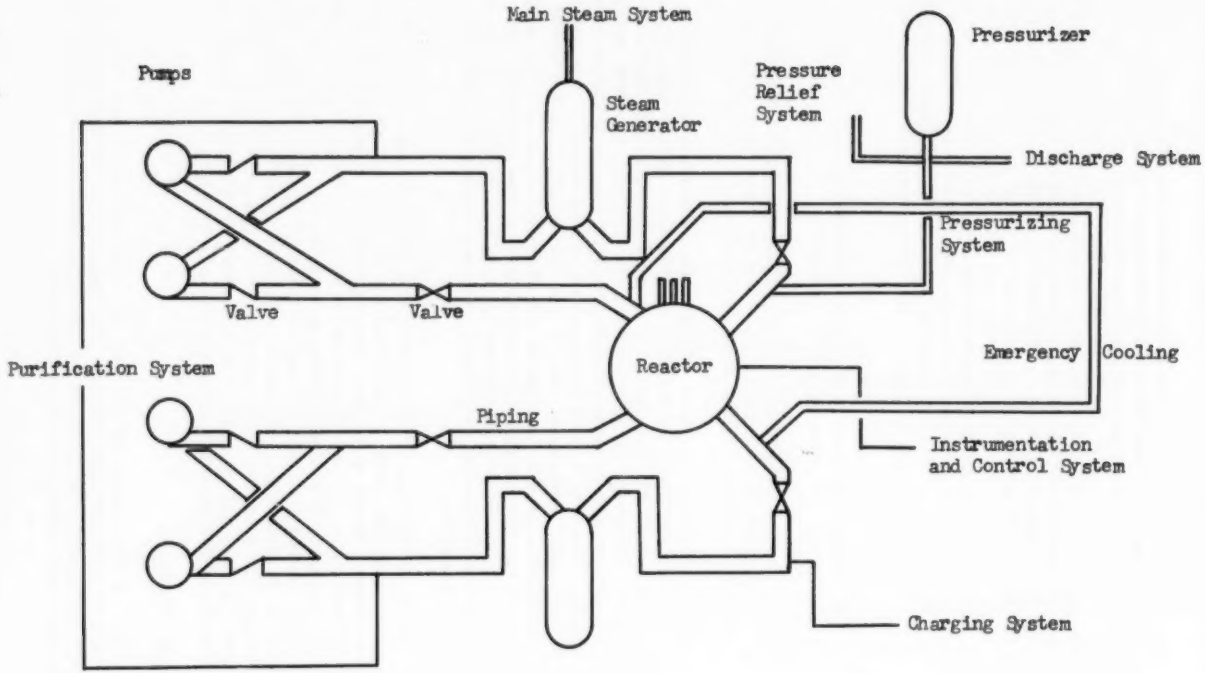
It is far more costly for both the Government and the equipment supplier when defective equipment is found at shipyard receiving inspection or in dockside trials after installation, than if found in the supplier's plant before shipment. A large amount of equipment has been returned to suppliers for repair because of defects found during shipyard receiving inspection. There have been numerous instances where it has been necessary to cut major components out of the system after installation and either conduct a repair program at the shipyard or return the equipment to the supplier for repair. The sea trials of one submarine were delayed several months because equipment quality was either clearly unsatisfactory or doubtful, and additional test and inspection was necessary to prove its integrity.

Before getting into this discussion, it might be well to show you just what a nuclear power plant consists of. A nuclear power plant, perhaps better termed a nuclear steam generator, takes the place of the conventional oil, gas, or coal fired boiler and consists of a mass of pressure vessels, heat exchangers, pumps, valves, controls, and interconnecting piping. Figure 1 is an outline of the main coolant system along with a designation of some of the more important auxiliary systems. The design pressure for this system is 2500 psi, the normal operating pressure is 2000 psi at a water temperature of about 500°F, and the primary construction material is 304 stainless steel.

Quality Problems

Most of this paper is devoted to a discussion of the various methods used to minimize and overcome the quality problems encountered in this program, but I would like to highlight some of the problem areas which have made remedial action necessary.

Figure 1.
Outline of Two Loop Pressurized Water Nuclear Power Plant



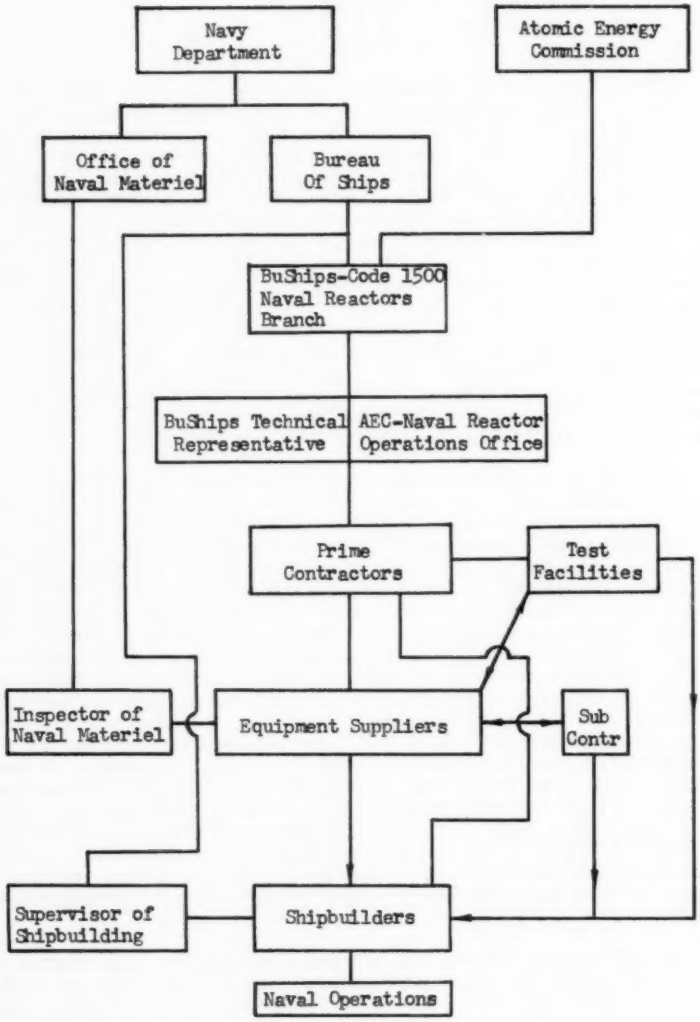
1. High quality requirements
2. Qualitative testing requirements
3. The maze of agencies involved
4. Supplier failures
5. Training

The overall quality requirements for a nuclear power plant are very high. Plant safety and integrity are of first order importance and no matter what the urgency of the situation from a schedular point of view, plant integrity is never compromised. The nuclear power plant has limited accessibility for maintenance because of the radiation hazards and because the system is all welded construction and therefore the operating reliability of the equipment must be exceptionally high. There are still a large number of operational and environmental unknowns and uncertainties, and these require that safety factors be used which are somewhat higher than normal. Conservatism of design is essential. The United States has gained a great amount of international technical prestige by virtue of the unprecedented success of the first nuclear powered submarines. The prestige could have been negative. High quality requirements and an insistence that these requirements not be compromised assured success of the Navy program. All of these factors keep quality requirements high and these are difficult to reduce unless there is ample evidence justifying the reduction in requirements.

Qualitative requirements are always a plague to Quality Control personnel and nuclear power plants have their share of these requirements. Basic equipment specifications are definitive and no problem exists in knowing what these specifications mean, but several of the most important tests used to evaluate equipment integrity are completely or partially qualitative. For example, the specification for a valve may include the requirement that the body be a class II, 304 stainless steel casting. This requirement leaves no question as to what is meant, but the determination of whether or not a casting meets class II requirements is sometimes very difficult. Our mechanical equipment has requirements for cleanliness and surface finish, and has nondestructive testing requirements in the areas of magnetic particle, liquid penetrant, ultrasonics, and radiography. These tests are interpretative in nature and compliance with requirements is often difficult to determine. In the case of ultrasonics, techniques are new and no standards have been established so that the test is used for information purposes only with defect indications explored by other methods. In the case of radiography, normal commercial techniques and results have been totally inadequate, standards are visual and comparative, and there is a very high premium on the experience factor both in technique and interpretation of results.

The relationship of the many agencies involved in this program is complex and unique and these relationships create communications problems which affect quality. The prime contractors are engineering and procurement facilities only, and they actually manufacture no part of the power plant itself. The various parts of the power plant are made under contract by suppliers from coast to coast and the plant is assembled and tested by the shipbuilder - either Navy or industrial - and the finished product is then operated by Naval personnel. Figure 2 shows the relationship between these agencies and the various routes by which equipment can reach the shipyard.

Figure 2.



Arrows show flow of equipment

Another problem has been apathy on the part of suppliers who do not take time to understand contract requirements or who do not really believe these requirements are necessary and therefore make no real attempt to comply with them. They fail to realize they must not only produce a piece of equipment which meets contractual requirements, but that they must provide documentation to prove it. Some suppliers fail to provide any semblance of quality surveillance over their own operations, depending completely on prime contractor or Navy inspection. Some suppliers have major subcontractors, and in numerous cases equipment is shipped directly from subcontractors to construction sites. Failure of supplier's Engineering Departments to transmit equipment requirements to their own shop or to their subcontractors has been a continuing source of difficulty.

Training of supervision, operators, and inspectors in the specialized requirements for nuclear work has been slow and difficult. This, of course, ties in with other problems and particularly with the qualitative testing requirements and communication failures. Getting equipment requirements transmitted intact from the mind of the procurement engineer to the operator and inspector in the supplier's shop is extremely difficult.

Almost all equipment going into one of our nuclear power plants is shipped directly from supplier to construction site. This being the case, we must perform our inspection at the supplier's plant prior to shipment. The intensity of inspection effort varies from a final inspection only to complete inspection from start to finish of large, complex items. The latter type of inspection is performed only when we have an inspector in residence at the supplier's plant. The intensity of effort given to a particular piece of equipment is dependent on available manpower, the supplier, the importance of the equipment in the system, and a knowledge of trouble encountered with similar equipment in the past.

Remedial Action

At the present time, we perform a great deal of physical inspection since it is by this means that we determine whether an item already built meets specifications and should be shipped. We still require that all items shipped carry an OK tag from one of our inspectors. However, it is well known that physical inspection is only a short range solution to poor quality, and is perhaps the least effective of the measures used for control and prevention of defects. The primary responsibility for quality lies with the supplier and more specifically with the supplier's quality control. We have therefore devoted much time and effort to activities specifically designed to improve the probability that equipment will be made right the first time and to improving the quality control function of our suppliers so that we, as a customer, can rely on supplier quality control. Only by doing this can we reduce our effort to surveillance and perform a very nominal amount of physical inspection.

The purpose of this paper is to describe these activities and to impress on those supplying equipment to the Naval Nuclear Power Program the necessity for exceptionally good quality control to implement the manufacture of nuclear power plant equipment.

Review of Specifications, Drawings, and Procedures

Before a contract is placed, our Quality Control Engineers review the specifications to check for equipment and testing requirements that are not clearly delineated and that the specification will be enforceable with a minimum of difficulty in the field. This action on our part prevents many difficulties before they start, as we attempt to eliminate and minimize the gray areas where the specification and requirements may not be clear. Most of our contracts specify that the supplier must submit drawings and various manufacturing procedures before he is released to proceed with fabrication. Most important of these are the welding, nondestructive testing, and functional testing procedures. Quality Control Engineers review these drawings and procedures before they are approved for use by the supplier. This review is done in the same manner as the original specification review. We make no attempt to pass on the adequacy of the design, leaving this area to the component engineer. Our primary concern is that the drawings and procedures accurately reflect the requirements of the contract and are satisfactory working documents for our field inspectors.

First Calls and Inspection Requirements

As soon as practical after an order is placed, we make a "first call" on the supplier to discuss order requirements and, in particular, to discuss "inspection requirements". Since Navy inspection is usually involved, the cognizant Navy inspector is invited to participate in this discussion. The time of the first call varies from several days after placement of an order to several months, depending on the type of order. When manufacturing on an order begins almost immediately after placement of the order, it is necessary that the first call be made very soon. On large complex equipment, some months may pass before the supplier's engineering and procurement of material are far enough along to where a discussion of quality and inspection is in order. First calls are by no means limited to new suppliers. First calls are made even more often on old suppliers receiving new contracts.

At the time of the first call, we try to make sure that the supplier understands the requirements of the order and especially the inspection and testing. We make sure that he understands what we require from him in the way of material certificates, logging of dimensions, nondestructive test results, and functional test data. We agree on a procedure by which he will notify us when he has reached a point where our inspection is required. We inform him of the procedure for obtaining deviations and in this connection we have attempted to make deviation approval more difficult to obtain. A supplier proposing to deliver equipment which deviates from requirements is required to submit a description and evaluation of the deficiencies in writing, stating that it is his belief that the deviation, if approved, will not affect the function, safety, reliability, guarantee, or service life of the apparatus to be delivered. Sketches and supporting calculations may be necessary to accurately describe or evaluate the condition.

After the first call is made, Inspection Requirements are written and transmitted to the vendor outlining in as much detail as possible what inspection we will perform and what information the vendor must supply to us. Inspection Requirements are written on all orders, even when

a first call is not made.

Quality Control does everything possible to help the supplier start the order with complete understanding of the quality requirements and with every aid we can supply to make sure the order will be satisfactorily produced.

Development of New MIL Specifications

Most suppliers are supplying the same or similar equipment to more than one of the prime contractors. To a large degree, most of the prime contractors are using their own specifications in delineating equipment requirements and, although these specifications must be approved by Code 1500, they are always different in some details and these differences have been a source of difficulty for suppliers. For instance, although the cleanliness requirements for all nuclear equipment is essentially the same, each of the prime contractors had his own cleanliness specification and they were different in some details so that what might satisfy one prime contractor would not satisfy another and this made for a difficult situation for the equipment supplier.

In order to minimize this problem, Code 1500 and the various prime contractors have initiated a program to put all the general requirements for nuclear equipment into MIL specifications and to use these MIL specs in lieu of individual company specifications. Much progress has been made in this area and such specifications as MIL-C-19874, Cleaning; MIL-STD-271, Nondestructive Test Requirements for Metals; MIL-STD-278, Welding; and MIL-STD-700, Locking Devices are the results of this program. This has been a great help to suppliers because at least they can use the same specification in dealing with the various customers. In the area of radiography, for instance, the use of MIL-STD-271 eliminates some of the variations in permissible techniques which previously existed. Along the same lines, MIL specifications are being written for the various pieces of equipment which go into the nuclear system and these MIL specs outline general equipment requirements which are common to all units of a particular type. These MIL specs are flexible in that there is an ordering data section in which the engineer specifies particular requirements of size, flow, materials, etc. For example, there is a MIL spec for globe valves from 1/2 to 4".

This trend toward the use of standard MIL specifications has not limited prime contractors in their ability to spell out equipment requirements, but rather has made them fit their requirements into the framework of the existing MIL specifications.

Special Instructions to Suppliers

Occasions sometimes arise which make it prudent to issue special instructions and information to all suppliers. Until recently, specifications have always required 100% radiography of large stainless steel castings. To the uninitiated, the term 100% is about as well defined as anything can be. However, in actual practice, liberties are sometimes taken even with requirements as well defined as this. We therefore found it necessary to send a letter to all of our casting suppliers and the following is an excerpt from this letter:

"Our interpretation of complete radiographic coverage is that all of the volume of metal in the casting be radiographed to the sensitivity required. More specifically, this means 100% of every cubic inch of every portion of every casting unless formally exempt by specific approval in line with the "identified series" approval of MIL-S-18262 (7 February 1958) or otherwise."

A similar situation arose concerning welds and it was necessary to write a special letter to all of our suppliers spelling out in considerable detail some special procedural requirements which we felt would be of mutual benefit to our suppliers and ourselves.

After some locking devices failed, instructions were issued to our suppliers and to our field inspectors that the securing of all locking devices was a mandatory check point which had to be witnessed by our inspectors. We made it clear to our suppliers that should they fail to inform us when locking operations were performed, that they would be required to disassemble and repeat the operation so that it could be witnessed.

In no case have we asked for more than the specifications called for but we did define in detail our interpretation of general requirements. We found suppliers generally cooperative in this effort since requirements were more clearly spelled out, and their opportunities for error were considerably reduced.

Special Informational Effort

Because of the difficulties which arose in the interpretation and usage of the qualitative testing requirements, it was necessary that some attempt be made to get all the prime contractors, suppliers, and shipyards to have a common understanding of these qualitative requirements. To implement this, a group was formed consisting of a Naval officer, Materials Engineering representatives from two prime contractors, and Quality Control representatives from two prime contractors. This group visited most of the major suppliers and all of the shipyards. At each supplier location, all of the supervision directly concerned with nuclear work was instructed and informed on the difficult qualitative specifications such as cleaning, welding, and nondestructive testing. This instruction was directed toward defining the intent of the specifications for these people, and answering any questions which they had concerning interpretation of the details of the requirements. It was usually found that when people understood the philosophy and intent of the requirements, they were much better equipped to deal with the details.

This effort was successful in eliminating some of the troublesome areas and in getting the various agencies to use and interpret the same specification in the same way.

Movie

Since each supplier is normally associated with a relatively small part of the system and since some of the requirements of the equipment which they supply are related to other more general aspects of the system, suppliers are prone to discount the necessity for some of the stringent quality requirements. Most Quality Control personnel are familiar

with the fact that if the man performing the job is aware of the requirements and convinced of the necessity for these requirements, he will always try to do a good job. In this program, the individual operators and inspectors in supplier's plants are extremely far removed from any concept of the overall system and system requirements. For example, how can a welder or machinist in a plant know that chips, slag, or dirt of rather small size can get into the small passages between fuel elements and produce a hot spot which might cause premature failure of a fuel element?

In order to bridge this gap, a thirty minute color movie is being made and will be shown to all suppliers. This movie is directed primarily at the operator and first line supervisor in order to provide them with a better understanding of nuclear systems and the reasons for some of the stringent requirements. It is anticipated that this movie will provide motivation to this large mass of individuals to improve the workmanship on nuclear equipment.

Trouble Reports

When equipment is found defective at the shipyard, a report is written by the yard describing the deficiency and the action taken to correct it. We investigate each of these reports to determine the reason for the deficiency and to determine whether our field inspectors could or should have prevented the incident. These investigations very often have proven fruitful in that they show up areas of weakness in our inspection operations. When this occurs, the incident is written up and distributed to all of our field personnel and to INSMAT, the Navy inspection service. These trouble reports are primarily instructional in nature and they point out to inspection personnel actual incidents that have happened and the corrective action which they need to take to prevent repetition of similar incidents. Criticism of individual inspectors is avoided in these Trouble Reports.

Recently several seams of two (2) pressure vessels were welded using the wrong weld wire. This was an automatic welding process and it is within the inspector's scope of responsibility to be sure that the weld wire used is correct. However, in practice, this is very difficult unless the inspector actually watches the wire drawn from stock and put on the automatic welding machine. After this incident occurred and it was determined how it happened, the supplier took steps to improve his control and a Trouble Report was written to all our field inspectors and to the Navy inspectors alerting them to the possibilities of such an error. This resulted in showing the same weakness to be present in other plants and a general improvement in control of welding wire.

Evaluation of Suppliers for Quality Control

The number of suppliers to the Naval Nuclear Propulsion Program has been limited for several reasons, among which are the size and complexity of equipment, the stringent requirements, the large amount of engineering talent necessary to design and build equipment for nuclear work, long lead times, heavy original investment on the part of suppliers, and a normal reluctance on the part of any procurement agency to have more suppliers than are really needed. With some notable exceptions, the suppliers to this program have had Quality Control organizations that were

rather weak. Some suppliers had no Quality Control organization per se and some suppliers have had no Quality Control function in any form. Since the supplier has responsibility not only for manufacture but also for inspection of equipment and since we want to rely on the supplier's Quality Control, we have found it advantageous to make a real effort to improve the Quality Control Departments of our suppliers.

An effective method of accomplishing improvement is to send Quality Control Evaluation Teams into all of our major suppliers, whose purpose is to assess the quality control function of the supplier as related to the fabrication of nuclear equipment. These are normally five man teams consisting of a Naval officer, a team leader from one of the prime contractors, and one representative from each of the three major prime contractors. This team spends whatever time is necessary in the supplier's plant surveying and evaluating the various aspects of the supplier's quality control function. This evaluation generally takes from one to four days.

After the evaluation is complete, a report is written to the company telling them of the areas which are considered weak and recommending changes to improve their internal quality control. This report is transmitted to the supplier by the Bureau of Ships and is, therefore, considered as official Navy evaluation and recommendations. The transmittal letter requests that the company review the report and reply to it, informing the Navy of what it intends to do relative to each of the recommendations. Three or four months after the evaluation, the Naval officer and team leader make a repeat visit to the supplier to check on the status of the supplier's compliance with the recommendation made.

These evaluations have been very helpful inasmuch as they have pinpointed areas of weakness for the suppliers, many of which the suppliers were not aware existed. Most of the recommendations have been accepted and implemented. Only in the area of organization changes have the suppliers been somewhat reluctant to carry out the recommendations, although some companies have made major organization changes as a result of the evaluation.

Training Program for Navy Inspectors

Although prime contractor field inspection, in addition to both supplier inspection and Navy inspection, is still necessary at this point in the program, it is not intended that this inspection continue indefinitely. However, prime contractor source inspection can only be suspended after supplier and Navy inspection prove to be reliable.

In order to assist the Navy inspection service in qualifying their personnel for nuclear work, a training program has been instituted. This program consists of sending specially selected Navy inspectors to a 4 or 6 week training school under our general direction and training them in the special requirements of nuclear work. These inspectors are picked on the basis of their education, background, and experience as being best suited for nuclear work. After the training course they will be permanently assigned to nuclear inspection work and normally given an increase in grade which will allow them to remain in nuclear work and be less susceptible to the normal transfers which take place.

Each of these schools consist of about 8 Navy inspectors and 3 or 4 contractor inspectors, with the contractor inspectors included on the basis that the association between Navy and prime contractor inspectors should be mutually beneficial. It is not expected that this training program will yield immediate or spectacular results, but in the long run it should ensure qualified nuclear component inspectors for the Navy inspection service.

Conclusion

Not all the above actions have been completely successful, but all have contributed something toward quality improvement of nuclear power plant equipment. Some of these actions are recent and others have been in effect for some time. There will most certainly be new problems arise which will require new and different actions. In any case, the lion's share of our long range planning and action is directed toward assurance that the equipment manufacturer will exercise the necessary degree of quality control so that we as a customer will have full confidence in his ability to supply good equipment to the program.

EXPERIMENTAL DESIGN - PART I
WHY DESIGN EXPERIMENTS?

Joseph Movshin
Ramsey Corporation

It takes very little contact with the American Industry to realize that a universal problem lies in the simple fact that there are always many factors which influence the results of even the simplest production process. The magnitude of this type of problem was first graphically demonstrated by F. W. Taylor* in his studies to standardize metal cutting activities. His specific question was, "What are the factors that determine metal cutting ability?" His goal was to find the combination of factors that would best produce the desired quality at the lowest cost for a given metal cutting problem. Ultimately these studies took 26 years, and eventually uncovered 12 factors which must be individually measured and controlled for the ultimate in metal cutting control.

Of course, all of these factors were not of the same magnitude. Many were of secondary importance. While they do contribute to the total effect, for all practical purposes, satisfactory results could be obtained by controlling a relatively few and standardizing many of the conditions governing the other factors. For example - on the average machine, tool feed and speed are variable as the primary controls. The choice of tool materials may be limited to relatively few in a given shop. In addition, the shape of the tool is standardized for a given kind of production operation. Probably one standard coolant is used on a given machine which covers a broader range of materials than theoretically might be desired. Obviously, inherent in the machine itself, are certain factors which are fixed by virtue of the fact that this particular machine is here and available.

But this problem, to a large degree, has been solved. We are interested in the innumerable other daily problems that occur. Changing circumstances have created the need to quickly and effectively identify sources of variation and to determine which of these are of sufficient magnitude to demand our attention and consideration. Unfortunately, this sort of problem as solved by Mr. Taylor, is extremely costly and time consuming. Perhaps in the more leisurely days at the end of the 19th century, we could have used Mr. Taylor's approach. This was to attempt to hold all the factors fixed with the exception of one which was varied to determine its effect. Then this is repeated for every other factor, and combinations of factors. Ultimately their relationships are developed.

* F. W. Taylor, Principles of Scientific Management (New York, Harper & Brothers Publishers, 1942) pp 104-109.

However, this is not fully adequate for many modern problems. There are complex interrelationships that would require simultaneous examination of several factors in various combinations. This kind of approach is obviously reserved for extreme laboratory situations, and even here we can offer statistical help with the techniques we are to discuss in this series.

The thing we are most interested in studying is variation. Variation occurs betwen successive tests involving what appears to be the same set of conditions. Variation also occurs between successive tests in which a factor is known to have changed. Did the change in the factor cause the variation in this case? Differences will occur between successive tests where it is not known if a factor has changed or not. How can we know that any specific change in a factor actually produces an essentially different value in the characteristic measured?

The most generally used statistical measurement of variation is the familiar standard deviation, or its square, the variance. This is used because it is the most effective statistic for measuring and comparing variations.

The basic analysis of variance in principle, is very simple. If we arrange a series of measurements from a given source in rows and columns, we can obtain several estimates of the variation inherent in the source. These are:

1) We can consider each column as a sub-group. From the between-column variation we arrive at one estimate of the total variation of the source.

2) We can consider each row as a sub-group. From the between-row variation we arrive at a second estimate of the total variation of the source.

3) Considering all the data as a single sub-group, we can make still another estimate of the variation of the source.

4) For each individual type of data arrangement, we can consider diagonals or other groupings as specific sub-groups. These in turn can lead to additional estimates of variation of the source.

We assume all the data is representative of the same source with no particular factor influencing any of the sub-groups. Then all of the estimates of the total source standard deviation must be compatible with each other. That is, they must be in statistical agreement subject only to chance variation.

The following table shows an arrangement of four rows and five columns. The rows might represent four operators (1, 2, 3, & 4) performing on the five machine (A, B, C, D, & E). Each number in the body of the table is a measurement of product or performance for a given operator and machine combination. For example, operator No. 2 on machine "C" produced the value 16. It might seem from a superficial examination of the data that operator No. 3 produces the highest value regardless of the machines he is using, while operator No. 1 has the lowest. Also machines A, B, C, appear different (produce a higher value) than D & E. However, if an analysis of variance were performed no significant difference between either the operators or machines would be found.

"Significance" in such proglems is measured by the "F" ratio. The "F" ratio for the rows, or operators was 1.73. This shows considerably more than a 10% probability that chance alone produced the variation seen. The "F" ratio for the columns was 3.92. There is also a much greater probability than 10% that chance could produce this value in this case. In both of these cases, therefore, we will assume there is no significance between the operators or between the machines. This is indeed fortunate because the numbers indicated in the table were obtained by jotting down a series of numbers from successive spins of a child's roulette wheel. These are random values and chance did produce the pattern seen.

		Columns - Machines					
		A	B	C	D	E	Row Totals
Rows - Operators	1	6	15	1	2	4	28
	2	18	8	16	7	3	52
	3	23	2	24	20	22	91
	4	0	20	7	7	3	37
Column Totals		47	45	48	36	32	208 Grand Total

We can add an assignable cause to one of the rows by the addition of a fixed amount (say 20) to each value in the row. The analysis of variance would then show a significant difference between rows, but no significant difference between columns. In other words, the "F" ratio would show that the variation between the columns was still that which could be produced by chance. However, the variation between the rows would be greater than we would expect to be produced by chance from the overall pattern of variation which produced the other three rows,

Without going into too much detail, the "F" ratio is a statistical device whereby the significance of the observed difference between two variances can be judged. Both variances represent an estimate of the variance of the assumed common source. If we divide the larger by the smaller of the two variances, the resulting number is the "F" ratio.

From a table we can determine the probability associated with chance alone giving rise to the observed "F" ratio. * For general statistical purposes we say that a 5% level of probability is "significant". This means that we assume that if chance should produce the observed difference only 5% of the time, we would be inclined to say that chance did not do so in the observed case. In the same way the 1% level is considered "highly significant". If the observed ratio is so large that there is only a 1% probability of chance producing it, we would quite readily accept the idea that chance did not produce the observed difference. We would assume that some highly significant factor actually accounted for the observed difference.

In the example above, the estimates of standard deviation from the rows and from the columns were each compared with the estimate of standard deviation based upon all the individual measurements. Statistically, this is a check for evidence of a significant source of variation in the factor representing the rows, or in the factor representing the columns. In problems of this sort, it is important to clearly define certain categories of information. First, we must ask, "What do we know".

1) Generally speaking, this question will be in the form of, "Does factor "A" have a significant effect upon the variation observed when the several possible sources of values of factor "A" are considered?". We may also do this for factor "B" along with factor "A". We may consider factor "C" in addition, or possibly ask, "Is there evidence of other factors which we are not considering?". We may also ask, "Is there interaction?". This refers to a combination of factors, the effect of which is not necessarily detected in the factors themselves.

The best description I know of interaction is the relationship between noise and age of children. If other children visit our children, it seems that as our children get older, the total noise level in the house increases. Suddenly, however, these visits which now take place in the evening, result in quiet. This happened about when our daughter turned 16. When her boy friends come over, we have an interaction that we did not have when our daughter was younger. I might point out that for my daughter alone, the rising noise level with age, still holds true in the day time.

* See - A. J. Duncan, Quality Control and Industrial Statistics (Chicago, Richard D. Irwin, Inc. 1952) pp 618-623. This table has "F" values for the probability of .5, .1, .05, .025, .01, .005, .001. Generally, tables have only the more commonly used values associated with a probability of .05 and .01.

Chart I
Interaction

Showing interaction of Day vs Night on the effect of age on noise level for daughters.

In addition to what we want to know, we must carefully consider what we do know.

1) What are the major contributing factors?

2) What are other possible contributing factors that we can readily identify?

3) What are the factors that can be randomized? These are the factors that occur throughout the major contributing sources with equal likelihood under any conditions. For example: Let us assume we have one source of steel which we are using in a process but run it on two machines and have a possible three operations. We might then assume that the steel factor is random, the machines appear to be one principle contributing factor, and the operators another contributing factor.

4) What are the constant factors? These are items that do not change.

4) (Cont'd)

In the illustration above, we might assume that the temperature and humidity in our building is essentially constant since the variation due to these is small enough to be assumed to be negligible.

In any of the above assumptions we must be alert to the fact that what we think is a contributing factor may contribute little to the variation. Also what we think is random, may not be. What we think is a negligible factor may not be so. These, the analysis will determine. However, we must make preliminary assumptions in order to set up a statistical model for our analysis.

Perhaps we can see why we need designed experiments in an actual case history. This problem occurred in the development of a new component of a piston ring which was required for modern high-speed, high-compression engines. It is one part of a multiple piece steel oil control ring called a "segment". This "segment" is essentially a ribbon of steel coiled on edge to form a ring. However, this coiling must result in a shape so uniform that when placed in a cylindrical gage each point on the circumference of the ring makes contact with the surface of the gage. This characteristic of contact in older ring specifications was described in loose, general terms such as "full contact with only fringe light permitted", or "intermittent light extending not over X^{o} at any one point".

A system of point measurement was first developed since it became obvious that no comparative tests could be made unless a scale of measurement were provided. Once we have the scale of measurement we began to consider the nature of the information available. First, we had several suppliers of steel, mill A, B, and C. These were all quite reputable sources. The chemical and physical properties of their steel met our specifications. There were, however, feelings expressed in the shop about the rolling ability of one steel or another, and the tendency of one steel or another to form "dog legs". These were straight sections in the circumference of the ring causing objectionable "open light".

In addition to the three sources of steel, there were also four rolling machines of essentially the same design but made over a period of years as requirements increased. Each machine was attended by an operator. Of course, there were feelings in the shop that particular combinations of machines and operators were superior to others.

One basic question was, "Is the steel a factor in our ability to produce the desired shape in our segments?". Secondly, was the operator and machine combination a factor? Were there other factors?

We decided to consider the steel first. The steel was supplied in

rolls which are quite long. When once placed on the machine it was necessary to run the entire reel. However, steel had been supplied to the various machines and operators randomly in the past. While the steel identity was maintained, that of the operator and machine were not There were sampling results of the rings produced by each roll of steel in terms of the numerical scale of light. This is the basic information for a 1-way classification by steel sources. Data which represented all the machines, all the operators and the three steels was arranged by steel supplier as in Table I.

Table I

	A	B	C
	14	9	12
	13	11	10
	10	0	6
	6	13	20
	23	-6	15
	1	-5	23
	-1	5	13
	16	5	19
	16	-4	4
	6	13	13
Total	104	41	135

We are assuming that the machine and operator effects are randomly distributed in columns A, B, and C. Are there any differences in the steel sources? The analysis of variance which will be worked in a subsequent session indicated an "F" ratio of 5.06. A table for "F" values showed a value of 3.35 for a 5% probability level and 5.49 for a probability level of 1%. We then knew that the probability that chance alone can produce differences between the steels as great as that observed, (F = 5.06) was between 5% and 1%. On the basis of this, we decided that chance alone did not produce the observed result and that there was indication that the steel itself was a contributing factor to the problem of rolling good segments.

The next question considers the effect of machine-operator combinations and steel sources. A series of tests were made in which each of the three steels were randomly assigned in normal production to each machine-operator combination.

When a total of approximately five samples (rolls) had been run for each steel with each machine-operator combination, the following table was prepared. This table is essentially the same as the operator and machine arrangement discussed before.

Table II

Steel	Machine 1	2	3	4	Steel Totals
A	11.5	8.2	7.1	9.9	36.7
B	6.1	7.0	6.8	4.9	24.8
C	12.0	12.5	8.7	11.2	44.4
Machine Total	29.6	27.7	22.6	26.0	105.9

The detail of working this problem will be indicated later. However, it was found that there appeared to be no evidence of significance for machine-operator combinations. In other words, any of the machine-operator combinations were about the same as another However, the great effect of the steel again became apparent. Specifically the "F" value found for the effect of machine-operator combination was 1.3 compared to 4.76 for the 5% probability of chance producing the differences observed. For the effect of steels, however, an "F" of 10.74 was observed, compared to a value of 5.14 for the 5% probability and 10.92 for the 1% probability. For our purpose certainly we had enough information to justify further action. Additional studies were made which verified these and enabled us to select the better steels for segments rolling. The use of designed experiments enabled us to make this progress with minimum cost.

Referring to the randomized 1-way classification, we used a model which answered the question, "Does the steel have any significant effect on the ability to produce light-tight segments?". In this case the other factors mainly, operator-machine combinations, were randomized. The same assumption (randomization) was made, although not stated for all other possible factors.

In the case of the randomized 2-way classification we were interested in the effect of machine-operator combinations as well as that of steel. We could have a 2-way classification with replication. This means that we have more than one observation per cell in the block arrangement. There are other models, or other arrangements of data for analysis of variance. Some of these will be discussed in more detail in subsequent sessions in this series.

As stated earlier, there are certain fundamental items of information that must be carefully defined prior to the design of an experiment:

a) What do we want to know?

b) What information is known?

c) What is not known?

d) The criteria for evaluating results.

In the example indicated above, we had to specifically state our question. The basic information known was a quality measurement produced by various combinations of steel and operator-machine groupings. We did not know the effects of the steel nor did we know the effects of operator-machine combinations. We did not know whether there might be any other factors of appreciable effect whose existance was unknown to us. Using the "F" ratio, we selected as our criteria for evaluated results, the 5% and 1% level of probability as being "significant" and "highly significant" indications of effect.

However, while the above is necessary for the statistical selection and solution of the model, there is another very important part of this problem. That is, the economic problem.

Basically, we must evaluate the following:

a) What is the value of the information to be gained?

b) What will the data cost us?

1) Per test.
2) Per one model compared to another.
3) In terms of the risks.

In our particular case we were running a scrap rate of approximately 24%. This was approximately $4,000 per month. This included material for both ourselves and an industrial customer that we were interested in satisfying. We determined to evaluate customer satisfaction in terms of the scrap we would have to produce to get enough good rings to satisfy this account.

Actually, our preliminary tests in this area required merely clerical work. Analysis of this original data cost us approximately $150. The next item to consider was how far could we afford to go in special tests to separate and correct sources of variation. Our preliminary check of past data indicated that the best scrap rate per lot that seemed to repeat periodically was 10%. We arbitrarily assumed the possibility of reaching this goal. This would mean a saving of $2300 per month. We then decided to make additional studies on more complicated statistical models than we have discussed above. The individual tests would cost about $180 each. This includes operator time, material and overhead. It would require approximately 20 test runs in the first series.

Management was approached with the idea of spending $3600 on the basis of a possible saving of $2300 per month, if our original guesses were right. This looked like a good gamble and Management took it.

The first series was to be a group of 2-way classifications in which certain steels would be selected for running on various machine-operator combinations. These would be replicated three times and measure specifically the effect of each steel and the possible saving that could be realized by running one steel alone. This series was promising but not conclusive. One steel, if used exclusively on all machines, would give us an overall scrap of about 18%. This represented material gain over our previous average, but not enough. We still felt the ultimate potential of 10% was worth going after.

Another more complicated series of tests were set up in which certain characteristics of the steel were related to the final quality. This was a much more expensive series and required considerably more than the analysis of variance alone. Some correlation analyses were also made. In order to get the money for this series, we again pointed out to Management that we were talking about a total saving from the initial study of $2300 x 12 or $30,600, per year. Since we had already demonstrated that we could realize some of this savings, we were given the go-ahead. We also were required to give solemn assurance that we would stop spending money and interfering with production as soon as we were convinced it did not pay to go further.

In giving this assurance to Management we were saying that we would gamble on our statistics only where we felt that the odds were good enough for the required investment. This also means that we had to be reasonably sure as the work progresses that the cost of a wrong decision at any point would not be too much. For this reason we broke down our series of tests into increments so that the total cost of any one series would be about $2000. In that way, we could say that if we pursued an area (steel characteristic) which was actually a blind path, the total cost of this one series would not be excessive in terms of the potential gain. Such factors as steel hardness, percentages of various elements, dimensional variation, et cetera, were evaluated in such series.

Let us consider the cost of a wrong decision in general terms. Suppose our test was indetermined, that is, hovering just above the 5% level of probability. If, the data were relatively cheap and the amount of possible saving were appreciable, we would be justified in spending a reasonable amount to more accurately determine whether or not we did have enough data to answer a given question. Sometimes in order to do this, confidence limits are set to estimate the range in which the true value may lie. This will also be discussed in a subsequent session. In any event, our basic problem is to weigh, on one hand, the cost of getting enough data to adequately answer our specific question, against the possibility gain. Industry is rarely interested merely in statistical significance if it does not mean dollar savings.

In considering risks, we often ignore the cost of dropping a project.

We may do this when no significance is indicated mainly because of the relatively small amount of data. As this is an easy wrong decision to make, it is one we try to avoid. If the savings potential of a given investigation offers any promises at all, we will even assume significance at a 10% level on preliminary investigations to assure not discarding a truly significant factor because of not enough data. Of course, as subsequent tests reveal no significance, we must drop the project as soon as we are convinced of this fact.

This total series, when finished, took about one year. As a point of information, we still don't know what makes a steel roll into a good segment. However, the results of our tests and pressure on our suppliers in directions our tests indicated, has resulted in our getting steel consistantly which does produce better segments even though the exact source of this quality is not known to us. We can produce good segments with under 10% rolling scrap. The designed experiment did pay off.

Perhaps we can outline our approach to the question of the economic aspect of a designed analysis of variance.

A. Initial phase

1) Determine if a clear opportunity for saving seems present.

   a) Lower scrap.

   b) Lower operating costs.

   c) Better performance (in terms of dollar value per day or year of life).
   d) Better performance (in terms of savings in other areas).
   e) Lower material costs (yeald).

2) Make estimate of ultimate potential savings in dollars.

3) Estimate cost of first series of tests. These should offer the possibility of some profit in themselves.

4) Consider costs of an incorrect decision.

   a) Deciding that a factor is not significant when it is. (potential gain that will not be made).

   b) Deciding that a factor is significant when in fact, it is not (extra tests which will not result in any gain).

5) Decide upon acceptance criteria

5) (Cont'd)
   a) General values - 5% significant, 1% highly significant.

   b) If potential gain is great, increase probability value for significance (10% perhaps)
   c) If potential gain is small, use 5% level (if potential gain is very small, why bother at all?)

6) Conduct initial tests and indicate the results.

   a) Is this information that can be useful?

   b) Is more data needed?

   c) Can some savings be realized at once?

B. Detailed analysis phase.

This phase generally involves more detailed investigations, usually in several directions and often more sophisticated statistical models. It may also involve other statistical techniques. Generally speaking, this phase is entered into only after the initial phase clearly indicates the significance of certain factors or areas of investigation where the potential gain can justify the costs.

1) Develop clear statements of the questions.

2) Select mathematical models to answer one or more questions simultaneously.
3) Evaluate the costs of tests by the various models.

4) Estimate the possible relative order of the magnitude of the various factors. (use past data or "indicated guess" or "enginering opinion".)

5) Select the models and test programs which offer the possibility of getting the answers with the minimum cost of testing. (It is necessary to remember that for any given model, the amount of data required to indicate significance varies inversely with the relative magnitude of the effects of the factors being investigated. For example, it would require small amounts of data to detect a significant factor whose effect was many times greater than any other factor or combination of factors. On the other hand, if the relative magnitude of the factors are comparable or if appreciable interaction is involved, considerably greater amounts of data are required. Expert statistical guidance is often helpful here.)

6) Again, the acceptance criteria must be determined, balancing the possible gains against the possible costs. This includes:

   a) Possible gain from the isolation and measurement of significant factors.
   b) Possible cost in wasted testing where no significance is to be found.
   c) Possible loss by assuming no significance and thereby not realizing the gain possible in isolating a significant factor.
   d) Possible cost of unnecessary testing and perhaps processing, in assuming significance in the factor where, in fact, none exists.

C. Final Phase -- Verification

   1) We now have enough detailed analysis to determine

      a) What factor can be acted upon and how
      b) What factors should not be considered at this time.
      c) What the ultimate goal in terms of product characteristic is to be.

   2) Once the changes are installed, a simple test should be arranged to check the following:

      a) That the total effect on the product is as desired.
      b) How the actual savings realized, compare with the original estimate of potential savings. (This last item helps develop the ability to do better on original estimates for subsequent tests.)

The above outline leaves a great deal to what might be called, "engineeing and managerial judgment?. For this we have no statistical substitute. The gambles worthwhile for the costs involved in a given situation, can only be decided by a management decision considering all the economic factors of the company, product, and future potential. The feasibility of engineering change or control of factors is one that can be answered only by imaginative engineering and design personnel who may be approached with a specific question, "can we do this with respect to a given factor?".

Looking into the future as we are doing in many of these instances, we can only use statistics to provide indication of the potential and the opportunities. Management and engineering must decide which gambles are worthwhile and what is practical. The actual statistical inferences which can be made from designed experiments are only guides to help point the way.

This introduction to designed analysis of variance has stressed the non-statistical aspects of the use of this technique. These administrative considerations are rarely discussed in statistical texts. Nevertheless, they are vitally necessary to the successful application of this tool in the industrial situation. It is for this reason that we have chosen to devote so much time to discussing this problem.

The subsequent papers of this series will deal with the actual solution of the statistical models. They will generally pre-suppose that the administrative decisions referred to in this paper will have been made prior to the actual solution of the analysis of variance problem itself.

EXPERIMENTAL DESIGNS AND ANALYSIS OF VARIANCE

II. ANALYSIS OF THE DATA

Max Astrachan
Air Force Institute of Technology

1. Introduction

In the first paper in this series we have been introduced to problems in which it is desired to analyze the variation arising from several possible sources. Experimental designs have been presented which can be used in different situations depending upon the number of sources of variation we can and want to analyze, and possible interrelationships among the sources.

In this paper we shall discuss in some detail two of the designs introduced in Part I of this series: one-way classification (completely randomized design, or one-factor experiments) and two-way classification (randomized block design, or two-factor experiments) with one observation for each combination of levels of the two factors.

Using a set of simple data, we shall show how to make the necessary calculations and tests, discuss the assumptions made and the implications of the results. We then apply the procedures to some of the data introduced in the examples of the first paper.

2. Basic Computations

Consider the following data:

Table 1

	A	B	C	D	
	8	10	18	7	
	1	6	15	1	
	3	11	12	16	
Total	12	27	45	24	108
Mean	4	9	15	8	9

Although the data are fictitious, the numbers could be thought of, for example, as the tensile strengths in certain units of a particular type of casting, with the figures representing the test results on 12 castings, three from each of four different suppliers, A, B, C, and D. Our problem is to decide whether the castings coming from different sources differ significantly with respect to their average tensile strengths. In other words, is the variation among the column means, 4, 9, 15, and 8, greater than one would expect to get by pure chance? We shall answer this question by analyzing variances.

Let us recall that for a set of N observations, X_1, X_2, . . . X_N, the variance of the X's is defined as

$$s^2 = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N - 1}$$

where $\bar{X}$ is the mean of the X's and N - 1 is the number of degrees of freedom. The numerator is usually referred to as sum of squares and is more conveniently calculated, especially when the number of observations is large or the mean is a decimal, from the formula

$$\sum_{i=1}^{N} (X_i - \bar{X})^2 = \sum_{i=1}^{N} X_i^2 - \frac{\left(\sum_{i=1}^{N} X_i\right)^2}{N} \qquad (1)$$

Now it can be proved algebraically* that the total variation in the data is equal to the variation within columns plus the variation of the column means about the grand mean. More specifically, the sum of squares of all the observations about their mean is equal to the sum of squares of the values in the columns about their respective means, plus the weighted sum of squares of the column means about their mean (the mean of all the observations). The weights are the numbers of observations in the columns.

We have then, for the total sum of squares,

$$(8 - 9)^2 + (1 - 9)^2 + (3 - 9)^2 + (10 - 9)^2 + \ldots + (16 - 9)^2$$

$$= (-1)^2 + (-8)^2 + (-6)^2 + (1)^2 + \ldots + (7)^2$$

$$= 1 + 64 + 36 + 1 + \ldots + 49 = 358.$$

For the columns, the sums of squares are

Col A: $(8 - 4)^2 + (1 - 4)^2 + (3 - 4)^2 = 16 + 9 + 1 = 26$

Col B: $(10 - 9)^2 + (6 - 9)^2 + (11 - 9)^2 = 1 + 9 + 4 = 14$

Col C: $(18 - 15)^2 + (15 - 15)^2 + (12 - 15)^2 = 9 + 0 + 9 = 18$

Col D: $(7 - 8)^2 + (1 - 8)^2 + (16 - 8)^2 = 1 + 49 + 64 = 114$

giving a total of 172.

---

* See Section 4.

For the weighted sum of squares for means, we have

$$3(4 - 9)^2 + 3(9 - 9)^2 + 3(15 - 9)^2 + 3(8 - 9)^2$$

$$= 3(25 + 0 + 36 + 1) = 3(62) = 186$$

Since 172 + 186 = 358, the assertion made earlier is verified.

To find these values by use of formula (1), we proceed as follows:

$$\text{Total sum of squares} = 8^2 + 1^2 + 3^2 + \ldots + 16^2 - \frac{(108)^2}{12}$$

$$= 1330 - 972 = 358$$

$$\text{Sum of squares for means} = 3(4^2 + 9^2 + 15^2 + 8^2) - \frac{(108)^2}{12}$$

$$= 3(16 + 81 + 225 + 64) - 972$$

$$= 1158 - 972 = 186$$

More conveniently, this last value can be computed by working with column totals and using the factor 1/3 instead of 3:

$$\frac{1}{3}(12^2 + 27^2 + 45^2 + 24^2) - \frac{(108)^2}{12} = \frac{3474}{3} - 972 = 1158 - 972 = 186$$

$$\text{Sum of squares within columns} = (8^2 + 1^2 + 3^2) - \frac{12^2}{3}$$

$$+ (10^2 + 6^2 + 11^2) - \frac{27^2}{3} + (18^2 + 15^2 + 12^2) - \frac{45^2}{3}$$

$$+ (7^2 + 1^2 + 16^2) - \frac{24^2}{3} = 8^2 + 1^2 + 3^2 + 10^2 + \ldots$$

$$+ 16^2 - \frac{1}{3}(12^2 + 27^2 + 45^2 + 16^2) = 1330 - 1158 = 172$$

In practice, we obtain the within-column sum of squares by subtraction: 358 - 186 = 172.

3. Testing for Significance

Our problem is to determine whether or not the column means differ significantly, i.e., whether the suppliers are providing us with castings of equal average tensile strengths. Our assumptions are that the readings in each column are from normal populations with equal variances.* This means that we are assuming that

* This can be tested by using Bartlett's test for homogeneity of variances. See Reference [2], p. 179.

$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma^2$, the common population variance. The hypothesis we are testing is $H_o: \mu_1 = \mu_2 = \mu_3 = \mu_4$.

In order to do this we shall obtain estimates of σ^2. One such estimate is found by dividing the total sum of squares, 358, by the number of degrees of freedom, 12 - 1 = 11. From each of the column sum of squares we can also obtain an estimate of σ^2 by dividing by 3 - 1 = 2 degrees of freedom. However, a better estimate can be obtained by pooling these four results using the formula

$$s_p^2 = \frac{(N_1 - 1)\, s_1^2 + (N_2 - 1)\, s_2^2 + (N_3 - 1)\, s_3^2 + (N_4 - 1)\, s_4^2}{\sum_{i=1}^{4} N_i - 4}$$

to give

$$s_p^2 = \frac{26 + 14 + 18 + 114}{8} = \frac{172}{8} = 21.5$$

A third estimate can be obtained by dividing the sum of squares for column means by 4 - 1 = 3 degrees of freedom. Each of these estimates of σ^2 is called a mean square. The second and third are of greatest interest. The results are summarized in the following analysis of variance table:

Table 2

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square
Means	186	3	62.0
Within	172	8	21.5
Total	358	11	

If the hypothesis we are testing, that the column means are equal, is true, then the mean square for means should not be too different from the within-column mean square. If it is, then we would reject the hypothesis and conclude that there is a greater difference between the output of the four suppliers than can be attributed to chance.

The actual test of the hypothesis is formed by using the F-ratio,

$$F = \frac{62.0}{21.5} = 2.88,$$

the ratio of the two mean squares. The number of degrees of freedom associated with the numerator is 3 and with the denominator is 8. From the table of critical values given in the Appendix, we find that $F_{.95}(3,8) = 4.07$, and so the value of F from our data is not significant at the .05 level. Our conclusion, therefore, is that the four suppliers are sending us castings with the same average tensile strength.

4. Assumptions and Model for One-way Classification

In the general case there would be k columns (suppliers) and r observations (castings) in each column. These could be represented as follows:

Table 3

	X_{11}	X_{21}	.	.	.	X_{i1}	.	.	.	X_{k1}
	X_{12}	X_{22}	.	.	.	X_{i2}	.	.	.	X_{k2}
	.	.	.	.	.	.	.	.	.	.
	X_{1j}	X_{2j}	.	.	.	X_{ij}	.	.	.	X_{kj}
	.	.	.	.	.	.	.	.	.	.
	X_{1r}	X_{2r}	.	.	.	X_{ir}	.	.	.	X_{kr}
Total	T_{1+}	T_{2+}	.	.	.	T_{i+}	.	.	.	T_{k+}
Mean	$\bar{X}_{1+}$	$\bar{X}_{2+}$	.	.	.	$\bar{X}_{i+}$	.	.	.	$\bar{X}_{k+}$

T_{++} = Grand Total

$\bar{X}_{..}$ = Grand Mean

$$T_{i+} = X_{i1} + X_{i2} + \cdots + X_{ir} = \sum_{j=1}^{r} X_{ij}$$

$$T_{++} = T_{1+} + T_{2+} + \cdots + T_{k+} = \sum_{i=1}^{k} T_{i+} = \sum_{i=1}^{k} \sum_{j=1}^{r} X_{ij}$$

$$\bar{X}_{i+} = \frac{T_{i+}}{r} = \frac{1}{r} \sum_{j=1}^{r} X_{ij}$$

$$\bar{X}_{..} = \frac{1}{k} \sum_{i=1}^{k} \bar{X}_{i+} = \frac{1}{kr} \sum_{i=1}^{k} T_{i+} = \frac{1}{kr} \sum_{i=1}^{k} \sum_{j=1}^{r} X_{ij}$$

Here X_{ij} is the measurement (tensile strength) on the jth individual (casting) from the ith category (supplier). The sums and averages are computed as indicated and N = kr is the total number of observations.

Our procedure then assumes the following model or form for each observation:

$$X_{ij} = \mu + c_i + \epsilon_{ij}$$

where μ is the overall mean, c_i is a column factor which varies from column to column, and ϵ_{ij} is a random error. It can be shown that the best estimate of μ is $\bar{X}_{..}$, of c_i is $\bar{X}_{i+} - \bar{X}_{..}$, and of ϵ_{ij} is $X_{ij} - \bar{X}_{i+}$.

If in the identity

$$(X_{ij} - \bar{X}_{..}) = (X_{ij} - \bar{X}_{i+}) + (\bar{X}_{i+} - \bar{X}_{..})$$

we square both sides and sum first over all the observations within a category and then over all categories, we have

$$\sum_{i=1}^{k} \sum_{j=1}^{r} (X_{ij} - \bar{X}_{..})^2 = \sum_{i=1}^{k} \sum_{j=1}^{r} [(X_{ij} - \bar{X}_{i+}) + (\bar{X}_{i+} - \bar{X}_{..})]^2$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{r} (X_{ij} - \bar{X}_{i+})^2$$

$$+ 2 \sum_{i=1}^{k} \sum_{j=1}^{r} (X_{ij} - \bar{X}_{i+})(\bar{X}_{i+} - \bar{X}_{..})$$

$$+ \sum_{i=1}^{k} \sum_{j=1}^{r} (\bar{X}_{i+} - \bar{X}_{..})^2$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{r} (X_{ij} - \bar{X}_{i+}) + r \sum_{i=1}^{k} (X_{i+} - \bar{X}_{..})^2$$

The middle term on the right vanishes, since it can be written as

$$\sum_i \sum_j X_{ij} \bar{X}_{i+} - \sum_i \sum_j X_{ij} \bar{X}_{..} - \sum_i \sum_j \bar{X}_{i+} \bar{X}_{i+} + \sum_i \sum_j \bar{X}_{i+} \bar{X}_{..}$$

$$= \sum_i \bar{X}_{i+} \sum_j X_{ij} - \bar{X}_{..} \sum_i \sum_j X_{ij} - r \sum_i \bar{X}_{i+}^2 + r \bar{X}_{..} \sum_i \bar{X}_{i+}$$

$$= \sum_i \bar{X}_{i+} T_{i+} - rk\bar{X}_{..}^2 - r \sum_i \bar{X}_{i+}^2 + (r\bar{X}_{..})(k\bar{X}_{..})$$

$$= r \sum_i \bar{X}_{i+}^2 - rk\bar{X}_{..}^2 - r \sum_i \bar{X}_{i+} + rk\bar{X}_{..}^2$$

$$= 0$$

by use of the formulas below Table 3. Thus we have proved what was asserted earlier, viz., that the sum of the squares of the observations about their mean is equal to the weighted sum of squares for column means plus the sum of squares within columns.

The computational formulas can then be written as follows:*

Total sum of squares: $S_T = \sum_i \sum_j X_{ij}^2 - \frac{T_{++}^2}{kr}$

Sum of squares for means: $S_M = \frac{1}{r} \sum_i T_{i+}^2 - \frac{T_{++}^2}{kr}$

Within-column sum of squares: $S_W = \sum_i \sum_j X_{ij}^2 - \frac{1}{r} \sum_i T_{i+}^2$

The analysis of variance table is then set forth as in Table 4.

Table 4

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square
Mean	S_M	$k - 1$	$S_M/(k - 1)$
Within	S_W	$N - k$	$S_W/(N - k)$
Total	S_T	$N - 1$	

* The given formulas hold only if the number of observations in each column is the same. If the numbers in the columns are n_1, n_2, $\cdots$, n_k, respectively, and N is the total number of observations, the formulas become

$$S_T = \sum_i \sum_j X_{ij}^2 - \frac{T_{++}^2}{N}$$

$$S_M = \sum_i \frac{T_{i+}^2}{n_i} - \frac{T_{++}^2}{N}$$

$$S_W = \sum_i \sum_j X_{ij}^2 - \sum_i \frac{T_{i+}^2}{n_i}$$

The hypothesis we are testing is

$$H_o: \mu_1 = \mu_2 = \cdots = \mu_k,$$

i.e., the means of the k categories are equal.* The test statistic is

$$F = \frac{\text{Mean Square for Means}}{\text{Mean Square for Within Groups}} = \frac{S_M/(k-1)}{S_W/(N-k)}$$

If we assume that the observations are random samples from normal populations with equal variances, and that the hypothesis is true, the critical region is $F > F_{.95}(k-1, N-k)$ as given in the Appendix.

5. An Example from Industry

Let us turn now to the example of the subjective measurement of "light" as a characteristic of a piston ring, as discussed in the previous paper. There were three different sources for the steel used in the rings, and a sample of 10 rings from each yielded the following (coded) data:

Table 5

	Steel Source			
	A	B	C	
	14	9	12	
	13	11	10	
	10	0	6	
	6	13	20	
	23	-6	15	
	1	-5	23	
	-1	5	13	
	16	5	19	
	16	-4	4	
	6	13	13	
Total	104	41	135	280

Using the formulas of Section 4, we have then

$$S_T = 14^2 + 13^2 + 10^2 + \cdots + 4^2 + 13^2 - \frac{(280)^2}{30}$$

$$= 4296 - 2613.33 = 1682.67$$

* This could also be written as $H_o: c_1 = c_2 = \cdots = c_k = 0$, i.e., the column effects are zero.

$$S_M = \frac{1}{10}(104^2 + 41^2 + 135^2) - \frac{(280)^2}{30} = 3072.20 - 2613.33$$

$$= 458.87$$

$$S_W = S_T - S_M = 1682.67 - 458.87 = 1223.80$$

The analysis of variance table is

Table 6

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square
Means (Between steel sources)	458.87	2	229.44
Within	1223.80	27	45.33
Total	1682.67	29	

Our test statistic is $F = \frac{229.44}{45.33} = 5.06$, and the .05 critical value for 2 and 27 degrees of freedom is $F_{95}(2,27) = 3.35$. Since our value of 5.06 is greater than the critical value, we conclude that at the .05 level the steel source does have an effect on the subjective measurement of "light."

6. Two-way Classification

We consider next the case of a two-factor experiment with a single observation for each combination of levels of the two factors. This enables us to study the effect of two variables in a single experiment. Thus we might be interested in knowing whether or not there were differences in the tensile strengths of our castings due to different testing machines as well as suppliers. An experiment of this kind is sometimes called a factorial design. We could, of course, have more than one observation for each combination of levels. (This case is discussed in the next paper in this series.)

Consider the following set of fictitious data which have been selected for ease of computation:

Table 7

	A	B	C	D	Total	Mean
a	4	6	6	12	28	7
b	9	10	11	14	44	11
c	5	8	13	10	36	9
Total	18	24	30	36	108	
Mean	6	8	10	12		9

The computations are made in the same manner as for one variable of classification. However, in addition to finding a mean square for columns, we shall also compute a mean square for rows.

We have, then,

$$\text{Total sum of squares} = 4^2 + 9^2 + 5^2 + 6^2 + \ldots + 10 - \frac{108^2}{12}$$

$$= 1088 - 972 = 116$$

$$\text{Sum of squares for columns} = \frac{1}{3}(18^2 + 24^2 + 30^2 + 36^2) - \frac{108^2}{12}$$

$$= \frac{1}{3}(3096) - 972 = 1032 - 972 = 60$$

$$\text{Sum of squares for rows} = \frac{1}{4}(28^2 + 44^2 + 36^2) - \frac{108^2}{12}$$

$$= \frac{1}{4}(4016) - 972 = 1004 - 972 = 32$$

When the sums of squares for rows and columns are added together and subtracted from the total, we have 24 (116 - 60 - 32) left over. This is referred to as the residual sum of squares and is composed of interaction and error. Interaction is that part of the variation caused by interplay of the two factors. Thus there might be a tendency for the product of a certain supplier to test higher on a certain machine. Error is that part of the variation due to chance. Here we assume there is no interaction or that it is negligible so that all of our residual is due to error.

There is a method of calculating the residual directly, but it is easier and simpler to do it by subtraction. To proceed directly we adjust the measurements so that the row and column means are all equal to the grand mean 9. This involves adding +2, -2, 0 to the measurements in the three rows, respectively, and then adding +3, +1, -1, -2 to the respective column entries. The results are shown in the following table:

Table 7a

	A	B	C	D	Total	Mean
a	9	9	7	11	36	9
b	10	9	8	9	36	9
c	8	9	12	7	36	9
Total	27	27	27	27	108	
Mean	9	9	9	9		9

Since the row and column means are equal, all the variation in the observations is due to error. We have then, for error or residual

sum of squares,

$$(9 - 9)^2 + (10 - 9)^2 + (8 - 9)^2 + (9 - 9)^2 + \ldots + (7 - 9)^2$$

$$= 0 + 1 + 1 + 1 + \ldots + 4 = 24$$

as before.

Table 8 shows the analysis of variance procedure.

Table 8

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square
Row means	32	3 - 1 = 2	16
Column means	60	4 - 1 = 3	20
Residual	24	(3 - 1)(4 - 1) = 6	4
Total	116	11	

To test the row means for significance we compare the row mean square with the residual mean square:

$$F = \frac{16}{4} = 4.00$$

The critical value for 2 and 6 degrees of freedom at the .05 level of significance is $F_{95}(2,6) = 5.14$. Hence, we conclude there is no difference among the rows (testing machines). For column means, we find $F = \frac{20}{4} = 5$, and $F_{95}(3,6) = 4.76$, so the difference between columns (suppliers) is significant at the .05 level.

7. Assumptions and Model for Two-way Classification

In the general we shall have c values or levels (columns) for one variable, and r values or levels (rows) for the second variable arranged as in Table 9. The column sums and means are found as in Table 3. The row sums and means are found in a similar manner.

Our procedure assumes the following model:

$$X_{ij} = \mu + c_i + r_j + \epsilon_{ij}$$

where μ is the overall mean, c_i a column effect which varies from column to column, r_j is a row factor which varies from row to row,

Table 9

	First Variable					Total	Mean
	X_{11}	X_{21}	.	.	.	X_{c1}	T_{+1}	$\bar{X}_{.1}$
	X_{12}	X_{22}	.	.	.	X_{c2}	T_{+2}	$\bar{X}_{.2}$
	X_{13}	X_{23}	.	.	.	X_{c3}	T_{+3}	$\bar{X}_{.3}$
Second Variable	.	.	.	.	.	.	.	.
	X_{1r}	X_{2r}	.	.	.	X_{cr}	T_{+r}	$\bar{X}_{.r}$
Total	T_{1+}	T_{2+}	.	.	.	T_{c+}	T_{++}	
Mean	$\bar{X}_{1.}$	$\bar{X}_{2.}$	.	.	.	$\bar{X}_{c+}$		$\bar{X}_{..}$

and ϵ_{ij} is a random error. The various sums of squares are computed from the following formulas:

Total sum of squares: $S_T = \sum_i \sum_j X_{ij}^2 - \frac{T_{++}^2}{cr}$

Column sum of squares: $S_c = \frac{1}{c} \sum_j T_{+j}^2 - \frac{T_{++}^2}{cr}$

Row sum of squares: $S_r = \frac{1}{r} \sum_i T_{i+}^2 - \frac{T_{++}^2}{cr}$

Residual sum of squares: $S_R = S_T - (S_c + S_r)$

The analysis of variance table is shown in Table 10.

Table 10

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square
Column means	S_c	$c - 1$	$S_c/(c - 1)$
Row means	S_r	$r - 1$	$S_r/(r - 1)$
Residual	S_R	$(c - 1)(r - 1)$	$S_R/(c - 1)(r - 1)$
Total	S_T	$rc - 1$	

Again we assume that all the populations have equal variances. Our design enables us to test the hypothesis that the column means are all equal, i.e., H_o: $c_1 = c_2 = \ldots = c_r = 0$, independent of whether or not there are row effects. Similarly, we may test the hypothesis that the row means are all equal, i.e., H_o: $r_1 = r_2 = \ldots = r_c = 0$, independent of whether or not there are column effects.

The appropriate statistics to use are:

$$\text{For columns: } F = \frac{S_c}{c-1} \Big/ \frac{S_R}{(c-1)(r-1)}$$

$$\text{For rows: } F = \frac{S_r}{r-1} \Big/ \frac{S_R}{(c-1)(r-1)}$$

and these would be compared with the critical values (at the .05 level of significance),

$$F_{.95}(c-1, (c-1)(r-1)) \text{ and } F_{.95}(r-1, (c-1)(r-1))$$

respectively.

8. Example of a Two-way Classification

In the situation of the example in Section 5 it was decided to investigate simultaneously the effects of different steels and machine-operator combinations using three levels for the former and four levels for the latter. The data (coded) are exhibited in Table 11.

Table 11

	Machine-Operator Combinations				
Steel	1	2	3	4	Total
A	11.5	8.2	7.1	9.9	36.7
B	6.1	7.0	6.8	4.9	24.8
C	12.0	12.5	8.7	11.2	44.4
Total	29.6	27.7	22.6	26.0	105.9

Using the formulas of Section 7, we have the following:

$$S_T = 11.5^2 + 6.1^2 + 12.0^2 + 8.2^2 + \ldots + 11.2^2 - \frac{105.9^2}{12}$$

$$= 1005.75 - 934.56 = 71.19$$

$$S_o = \frac{1}{3}(29.6^2 + 27.7^2 + 22.6^2 + 26.0^2) - \frac{105.9^2}{12}$$

$$= 943.40 - 934.56 = 8.84$$

$$S_r = \frac{1}{4}(36.7^2 + 24.8^2 + 44.4^2) - \frac{105.9^2}{12}$$

$$= 983.32 - 934.56 = 48.76$$

$$S_R = 71.19 - (8.84 + 48.76)$$

$$= 71.19 - 57.60 = 13.59$$

The analysis of variance table can now be constructed as shown in Table 12.

Table 12

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square
Between Machine-Operators	8.84	3	2.95
Between Steels	48.76	2	24.38
Residual	13.59	6	2.27
Total	71.19	11	

For machine-operators,

$$F = \frac{2.95}{2.27}$$

and the critical value $F_{95}(3,6) = 4.76$. Hence, there is no difference between machine-operators.

For steels,

$$F = \frac{24.38}{2.27} = 10.74$$

and the critical value $F_{95}(2,6) = 5.14$. Hence, there is a significant difference between steels.

9. Multiple Comparisons

The basic test described in the analysis of the two models we have discussed determines whether or not there is a significant difference among a group of means. If the hypothesis of equal means is rejected, it cannot tell us which means differ significantly from which others. Often this is what the experimenter really wants to know. Further, he may want to establish confidence limits for such differing means, or he may want to compare two or more means taken from the entire set. In the case of acceptance of the hypothesis of equal means, he may want to establish confidence limits for the over-all population mean.

We shall omit the discussion of such questions from this paper because of lack of time and space. A complete discussion of these and other questions relating to multiple comparisons will be found in the two references given at the end of this paper.

References

1. Bennett, C. A., and Franklin, N. A. Statistical Analysis in Chemistry and the Chemical Industry, 1954. John Wiley and Sons, Inc., New York.

2. Dixon, W. J., and Massey, F. J., Jr. Introduction to Statistical Analysis, Second Edition, 1957. McGraw-Hill Book Company, Inc., New York.

APPENDIX

Percentiles of the $F(n_1, n_2)$ Distribution with n_1 Degrees of Freedom for the Numerator and n_2 for the Denominator.

n_2	P \ n_1	1	2	3	4	5	6	7	8	9	10	15	20	30	40	60	∞
1	.95	161	200	216	225	230	234	237	239	241	242	246	248	250	251	252	254
	.975	648	800	864	900	922	937	948	957	963	969	985	993	100^1	101^1	101^1	102^2
	.99	* 405^1	500^1	540^1	562^1	576^1	586^1	593^1	598^1	602^1	606^1	616^1	621^1	626^1	629^1	631^1	637^1
	.995	* 162^2	200^2	216^2	225^2	231^2	234^2	237^2	239^2	241^2	242^2	246^2	248^2	250^2	251^2	253^2	255^2
2	.95	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5
	.975	38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4	39.4	39.4	39.5	39.5	39.5	39.5
	.99	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5
	.995	198	199	199	199	199	199	199	199	199	199	199	199	199	199	199	200
3	.95	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.70	8.66	8.62	8.59	8.57	8.53
	.975	17.4	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.5	14.4	14.3	14.2	14.1	14.0	14.0	13.9
	.99	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	26.9	26.7	26.5	26.4	26.3	26.1
	.995	55.6	49.8	47.5	46.2	45.4	44.8	44.4	44.1	43.9	43.7	43.1	42.8	42.5	42.3	42.1	41.8
4	.95	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86	5.80	5.75	5.72	5.69	5.63
	.975	12.2	10.6	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.66	8.56	8.46	8.41	8.36	8.26
	.99	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.2	14.0	13.8	13.7	13.7	13.5
	.995	31.3	26.3	24.3	23.2	22.5	22.0	21.6	21.4	21.1	21.0	20.4	20.2	19.9	19.8	19.6	19.3
5	.95	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62	4.56	4.50	4.46	4.43	4.36
	.975	10.0	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.43	6.33	6.23	6.18	6.12	6.02
	.99	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.72	9.55	9.38	9.29	9.20	9.02
	.995	22.8	18.3	16.5	15.6	14.9	14.5	14.2	14.0	13.8	13.6	13.1	12.9	12.7	12.5	12.4	12.1

*See last page of Appendix for explanation.

*See last page of Appendix for explanation.

APPENDIX (Continued)

Percentiles of the $F(n_1, n_2)$ Distribution

n_2	P \ n_1	1	2	3	4	5	6	7	8	9	10	15	20	30	40	60	∞
6	.95	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94	3.87	3.81	3.77	3.74	3.67
	.975	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.27	5.17	5.07	5.01	4.96	4.85
	.99	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.56	7.40	7.23	7.14	7.06	6.88
	.995	18.6	14.5	12.9	12.0	11.5	11.1	10.8	10.6	10.4	10.2	9.81	9.59	9.36	9.24	9.12	8.88
7	.95	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.51	3.44	3.38	3.34	3.30	3.23
	.975	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.57	4.47	4.36	4.31	4.25	4.14
	.99	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.31	6.16	5.99	5.91	5.82	5.65
	.995	16.2	12.4	10.9	10.0	9.52	9.16	8.89	8.68	8.51	8.38	7.97	7.75	7.53	7.42	7.31	7.08
8	.95	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.22	3.15	3.08	3.04	3.01	2.93
	.975	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.10	4.00	3.89	3.84	3.78	3.67
	.99	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.52	5.36	5.20	5.12	5.03	4.86
	.995	14.7	11.0	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	6.81	6.61	6.40	6.29	6.18	5.95
9	.95	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.01	2.94	2.86	2.83	2.79	2.71
	.975	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.77	3.67	3.56	3.51	3.45	3.33
	.99	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	4.96	4.81	4.65	4.57	4.48	4.31
	.995	13.6	10.1	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	6.03	5.83	5.62	5.52	5.41	5.19
10	.95	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.85	2.77	2.70	2.66	2.62	2.54
	.975	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.52	3.42	3.31	3.26	3.20	3.08
	.99	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.56	4.41	4.25	4.17	4.08	3.91
	.995	12.8	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.47	5.27	5.07	4.97	4.86	4.64
15	.95	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.40	2.33	2.25	2.20	2.16	2.07
	.975	6.20	4.76	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.86	2.76	2.64	2.59	2.52	2.40
	.99	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.52	3.37	3.21	3.13	3.05	2.87
	.995	10.8	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.07	3.88	3.69	3.59	3.48	3.26

APPENDIX (Continued)

Percentiles of the $F(n_1, n_2)$ Distribution

n_2	P \ n_1	1	2	3	4	5	6	7	8	9	10	15	20	30	40	60	∞
20	.95	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.20	2.12	2.04	1.99	1.95	1.84
	.975	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.57	2.46	2.35	2.29	2.22	2.09
	.99	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.09	2.94	2.78	2.69	2.61	2.42
	.995	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	3.50	3.32	3.12	3.02	2.92	2.69
30	.95	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.01	1.93	1.84	1.79	1.74	1.62
	.975	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.31	2.20	2.07	2.01	1.94	1.79
	.99	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.70	2.55	2.39	2.30	2.21	2.01
	.995	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	3.01	2.82	2.63	2.52	2.42	2.18
40	.95	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.92	1.84	1.74	1.69	1.64	1.51
	.975	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.18	2.07	1.94	1.88	1.80	1.64
	.99	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.52	2.37	2.20	2.11	2.02	1.80
	.995	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22	3.12	2.78	2.60	2.40	2.30	2.18	1.93
60	.95	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.84	1.75	1.65	1.59	1.53	1.39
	.975	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.06	1.94	1.82	1.74	1.67	1.48
	.99	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.35	2.20	2.03	1.94	1.84	1.60
	.995	8.49	5.80	4.73	4.14	3.76	3.49	3.29	3.13	3.01	2.90	2.57	2.39	2.19	2.08	1.96	1.69
	.95	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.67	1.57	1.46	1.39	1.32	1.00
	.975	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.83	1.71	1.57	1.48	1.39	1.00
	.99	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.04	1.88	1.70	1.59	1.47	1.00
	.995	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62	2.52	2.19	2.00	1.79	1.67	1.53	1.00

1. The number 1 as a superscript indicates a product by 10
2. The number 2 as a superscript indicates a product by 100

Read 405^1 as 4050, and 162^2 as 16200.

EXPERIMENTAL DESIGN AND ANALYSIS OF VARIANCE - III
ADDITIONAL DESIGNS AND ANALYSES

Fred C. Leone
Case Institute of Technology

In the two preceding papers we presented some situations in which a designed experiment apply. We also noted the philosophy of a statistical design, the risks, the general method of attack, and the point of view of an individual in industry who wishes to obtain the maximum amount of information from his data. The collection of data without a proper design could be as bad as using a demi-tasse spoon to ladle soup. The results, (or lack of them) may not be as dramatic, but the efficiency could be comparable. In these papers we are discussing more than just designed experiments. We are considering statistically designed experiments and the analyses which follow.

In the earlier paper two experiments were discussed: (1) a single factor experiment, and (2) a two factor experiment with one observation per cell. We wish now to extend the latter to any number of observations per cell. Further, we shall discuss a crossed versus a nested (or hierarchal) experiment. An actual example of a two-factor crossed experiment will be discussed and analyzed. Finally, a Latin Square experiment will be presented by means of a brief example.

Consider a two factor experiment. Let us call these Factor A and Factor B. Factor A has p levels, and Factor B has q levels. What we would like to know is whether the different levels of Factor A produce a different average effect on the yield. We would also ask this question of Factor B. Finally, do the average effects change significantly for different combinations of levels of Factors A and B? We are here describing a two factor crossed experiment. The mathematical model of the yield (x) is as follows:

$$x_{ij\alpha} = \mu + e_{i.} + e_{.j} + e_{ij} + e_{\alpha(ij)}$$

with $i = 1, 2, \ldots, p$
$j = 1, 2, \ldots, q$
$\alpha = 1, 2, \ldots, n$
$pqn = N$

and μ = the grand mean

$e_{i.}$ = the effect of the i^{th} level of Factor A over and above μ

$e_{.j}$ = the effect of the j^{th} level of Factor B over and above μ

e_{ij} = the effect of the combination of the i^{th} and j^{th} levels respectively of A and B over and above μ, $e_{i.}$ and $e_{.j}$

$e_{\alpha(ij)}$ = the deviation from the expected value in a cell.

The theoretical array of means which the observed data represent are given in Table 1.

Table 1 TRUE MEANS OF A TWO FACTOR EXPERIMENT

		Levels of Factor A				
		1	2	...	p	B Means
Levels of FACTOR B	1	e_{11}	e_{21}	...	e_{p1}	$e_{.1}$
	2	e_{12}	e_{22}	...	e_{p2}	$e_{.2}$
		⋮	⋮	⋮	⋮	⋮
	q	e_{1q}	e_{2q}	...	e_{pq}	$e_{.q}$
A Means		$e_{1.}$	$e_{2.}$	...	$e_{p.}$	

The hypotheses which we are testing are:

(1) $e_{1.} = e_{2.} = \dots = e_{p.}$

(2) $e_{.1} = e_{.2} = \dots = e_{.q}$

(3) $e_{11} = e_{12} = \dots = e_{pq}$

In order to carry out these tests we obtain an unbiased estimate of σ^2, the within cell variance. (This is assumed to be constant within each cell). To test hypothesis (3) we obtain an estimate of the variance among interaction terms. By testing whether this is zero we are essentially testing whether each of the e_{ij} are equal. By a similar procedure we test for equality of average levels of the main effects designated in the first two hypotheses. To clarify, let us spell out the procedure.

(1) The Model is Specified.* This is part of the physical background of the experiment. For the present, let us assume a two factor crossed experiment with more than one observation per cell. (Consider an equal number of observations per cell). Assume also that the levels

* The importance of choice of the model should not be underestimated. The selection of the correct model, including the assumptions and such facts as levels of population on which inference is to be drawn, fixed or random factors, nested or crossed, quantitative or qualitative factors, additional tests of significance, confidence intervals and possible conclusions are important in the choice of the valid statistical tests and in the validity of the physical conclusions.

of both factors are random choices from a large number of choices. This model is, then,

$$x_{ij\alpha} = \mu + e_{i.} + e_{.j} + e_{ij} + e_{\alpha(ij)}$$

(2) The Data is Collected. This is, in general:

Table 2 DATA OF A TWO FACTOR EXPERIMENT

Factor B \ Factor A	1	2		p
1	x_{111}	x_{211}		x_{p11}
	x_{112}	x_{212}	...	x_{p12}
	⋮	⋮		⋮
	x_{11n}	x_{21n}		x_{p1n}
	...	...	$x_{ij\alpha}$	...
q	x_{1q1}	x_{2q1}		x_{pq1}
	x_{1q2}	x_{2q2}	...	x_{pq2}
	⋮	⋮		⋮
	x_{1qn}	x_{2qn}		x_{pqn}

Note here that we have n observations to estimate each of the true means of Table 1.

(3) Computation of Sum of Squares. We calculate the following:
(a) The cell totals, A totals, B totals.

Table 3 CELL AND MARGINAL TOTALS

Factor B \ Factor A	1	2	...	p	Total
1	T_{11}	T_{21}	...	T_{p1}	T_{+1}
2	T_{12}	T_{22}	...	T_{p2}	T_{+2}
	⋮	⋮	⋮	⋮	⋮
q	T_{1q}	T_{2q}	...	T_{pq}	T_{+q}
Total	T_{1+}	T_{2+}	...	T_{p+}	T_{++}

(b) The Sums of Squares

$$CF^* = \frac{1}{N}(x_{111} + x_{112} + \ldots + x_{pqn})^2 = T_{++}^2/N$$

$$S_i = \frac{1}{qn}(T_{1+}^2 + T_{2+}^2 + \ldots + T_{p+}^2) - CF$$

$$= \frac{1}{qn}\sum_{i=1}^{p} T_{1+}^2 - CF$$

$$S_j = \frac{1}{pn}(T_{+1}^2 + T_{+2}^2 + \ldots + T_{+q}^2) - CF$$

$$= \frac{1}{pn}\sum_{j=1}^{q} T_{+j}^2 - CF$$

$$S_{ij} = \frac{1}{n}(T_{11}^2 + T_{12}^2 + \ldots + T_{pq}^2) - CF - S_i - S_j$$

$$= \frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{q} T_{ij}^2 - CF - S_i - S_j$$

$$S = (x_{111}^2 + x_{112}^2 + \ldots + x_{pqn}^2) - T_{++}^2/N$$

$$= \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{\beta=1}^{n} x_{ij\alpha}^2 - T_{++}^2/N$$

$$S_{\alpha(ij)} = S - S_i - S_j - S_{ij}$$

(4) The Analysis of Variance Table.

Table 4 ANALYSIS OF VARIANCE OF TWO FACTORS WITH REPLICATIONS

Source	Sum of Squares (S.S.)	Degrees of Freedom (d.f.)	Mean Square (M.S.)	Expected Value of
Factor A	S_i	p-1	$S_i/(p-1)$	$\sigma^2+n\sigma_{ij}^2+qn\sigma_i^2$
Factor B	S_j	q-1	$S_j/(q-1)$	$\sigma^2+n\sigma_{ij}^2+pn\sigma_j^2$
Interaction(AB)	S_{ij}	(p-1)(q-1)	$S_{ij}/(p-1)(q-1)$	$\sigma^2 + n\sigma_{ij}^2$
Residual	$S_{\alpha(ij)}$	pq(n-1)	$S_{\alpha(ij)}/pq(n-1)$	σ^2
Total	S	pqn-1 = N-1		

*This is sometimes designated the correction factor.

(5) The Tests of Significance. We must first test the hypothesis that the different combinations of the first and second factor do not produce significantly different effects on the yield. This is equivalent to asking whether these two factors could be considered as independent. We do this by taking the ratio

$$F_{ij} = \frac{S_{ij}/(p-1)(q-1)}{S_{\alpha(ij)}/pq(n-1)}$$

By this ratio we are testing whether σ_{ij}^2 is significantly different from zero. If σ_{ij}^2 is not significant then we accept the hypothesis that $e_{11} = e_{12} = \ldots = e_{pq}$.

This ratio is compared with the critical values of the F distribution $(F_{1-\alpha}, (p-1)(q-1), pq(n-1))$ for $\alpha = .05, .01$ and $.001$. If $F_{ij} < F_{.05}$ we accept the hypothesis of no difference. Otherwise we say that we have evidence of a real difference and reject the hypothesis.*

We consider next Factor B. In Table 4, by testing whether σ_j^2 is significantly different from zero, we are testing the hypothesis $e_{.1} = e_{.2} = \ldots = e_{.q}$. We likewise test Factor A.

$$F_j = \frac{S_j/(q-1)}{S_{\alpha(ij)}/pq(n-1)} \qquad \text{if } F_{ij} \text{ is not significant}^{**}$$

$$= \frac{S_j/(q-1)}{S_{ij}/(p-1)(q-1)} \qquad \text{if } F_{ij} \text{ is significant}$$

and $$F_i = \frac{S_i/(p-1)}{S_{\alpha(ij)}/pq(n-1)} \qquad \text{if } F_{ij} \text{ is not significant}$$

$$= \frac{S_i/(p-1)}{S_{ij}/(p-1)(q-1)} \qquad \text{if } F_{ij} \text{ is significant}$$

---

* A usual notation is the use of "*" for a significance level beyond 5 per cent but less than 1 per cent, "**" for a level beyond 1 per cent but less than 0.1 per cent and "***" for a level beyond 1 per cent.

** We shall not consider here the matter of pooling. This should be checked in the references given at the end of this paper.

(6) Conclusions. To simply carry out the tests of significance without any regard to further evaluation and the statement of these results in physical terms is an insult to the potential of this statistical technique and often to the individual who is presented with these results. It is like preparing an excellent meal and refusing to supply the equipment to properly handle the food.

It is not our purpose here to go into all of the additional statistical determinations and explain them in detail. Rather, let us simply state them in the hope that the reader may investigate these in the references cited and other books and articles. To simply say that there is, or is not, a significant difference among the levels of Factor A is not enough. We might ask, for example: (1) What are the estimates of the mean effect of each level? (2) Can we separate each of the levels into any particular grouping? (3) What about a $1-\alpha$ per cent confidence interval on the true mean of level i? (4) How large a difference do we need between the levels to say that this observed difference is real? (5) What are our best estimates of the different variances? (6) Just what have we learned from this experiment?

Let us now proceed with an actual example. This was part of an extensive problem on "Sandwich Structures" carried out in the Research and Development Department of the TAPCO Group of Thompson Ramo Wooldridge Inc. These "Sandwich Structures" are a honeycomb type of structure as shown in Figure 1. Four brazing materials (Factor A), namely, GE 81, AG-MN, Coast Metal 53, and Electrolytic Copper are to be evaluated as applied to the brazing of three honeycomb structure materials (Factor B), namely, 321 Stainless Steel, 17-7 Precipitation Hardening Steel, and Titanium. The one characteristic of the brazed joints considered in this portion of the project is uniformity of bond.

Three (3) samples of each combination of braze and structure materials are destructively examined by peeling the outer skin away from the core after brazing (see Figure 1). By visual and microscopic examinations, it may be determined whether or not a cell was completely brazed.

Each 4" x 8" panel contains 512 quarter-inch cells. We shall count the number of defective cells and record this as the yield, $x_{ij\alpha}$.

The definition of a defective cell would be: any cell not exhibiting complete braze filleting of the skin to the core around 100 per cent of its periphery.

We now proceed with the problem. Following the six steps as given earlier we have:

(1) The Model

$$x_{ij\alpha} = \mu + e_{i.} + e_{.j} + e_{ij} + e_{\alpha(ij)}$$

where $i = 1, 2, 3, 4$ $(= p)$ brazing materials

$j = 1, 2, 3$ $(= q)$ structure materials

$\alpha = 1, 2, 3,$ $(= n)$ samples for each combination of i and j.

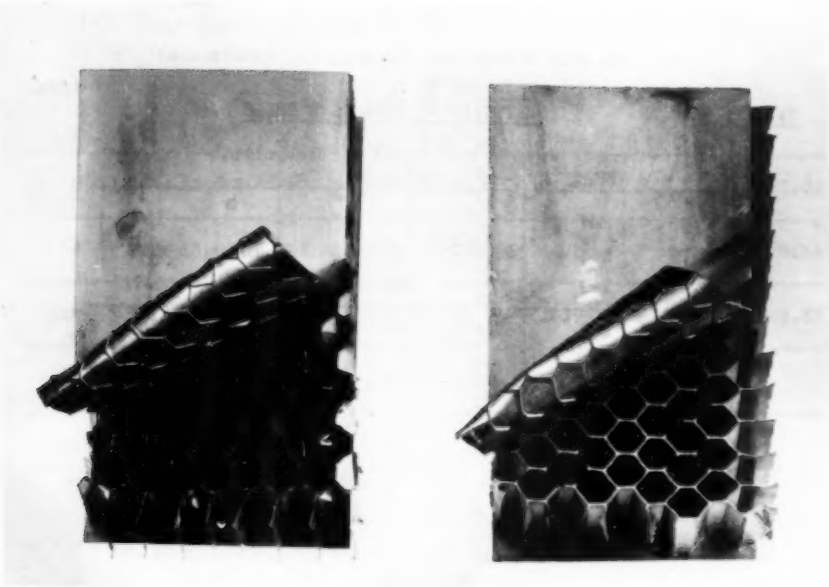


Figure 1.

In this example the conclusions are to be drawn only on the four brazing materials and three structure materials. Hence, the levels of these two factors are to be considered as fixed. This does not change any of the computational procedures except for the final F-ratios. This change will be noted.

(2) The Data. The results of the visual and microscopic examinations for defective cells are recorded.

Table 5 DEFECTIVE CELLS OF SANDWICH STRUCTURES

Structure Material	(Brazing Material) GE 81	Ag-Mn	Coast Metal 53	Electrolytic Copper
321 S.S.	35	2	0	4
	50	5	5	10
	22	2	1	6
17-7 PH	80	90	20	30
	85	110	10	20
	100	125	9	19
Titanium	50	40	40	27
	79	39	28	35
	60	60	55	26

(3) Computation of Sum of Squares.

(a) The cell totals and the marginal totals are:

Table 6 SUBTOTALS OF DEFECTIVE CELLS

	GE 81	Ag-Mn	Coast Metal 55	Electrolytic Copper	T_{+j}
321 S.S.	107	7	4	20	138
17-7 PH	265	325	39	69	698
Titanium	189	139	121	88	537
T_{i+}	561	471	164	177	1373 = T_{++}

(b) Sum of Squares.

$$CF = \frac{1}{36}(35 + 50 + \ldots + 26)^2 = 1373^2/36 = 52,364$$

$$S_i = \frac{1}{9}(561^2 + 471^2 + 164^2 + 177^2) - 1373^2/36$$

$$= 66,087 - 52,364 = \underline{13,723}$$

$$S_j = \frac{1}{12}(138^2 + 698^2 + 537^2) - 52,364$$

$$= 66,218 - 52,364 = \underline{13,854}$$

$$S_{ij} = \frac{1}{3}(107^2+265^2+\ldots+88^2) - 52,364 - 13,723 - 13,854$$

$$= 38,127 - 13,723 - 13,854 = \underline{10,550}$$

$$S = (35^2 + 50^2 + \ldots + 26^2) - 52,364 = \underline{40,709}$$

$$S_{a(ij)} = 40,709 - 13,723 - 13,854 - 10,550 = \underline{2,582}$$

(4) The Analysis of Variance Table and (5) Tests of Significance.

Table 7 ANALYSIS OF VARIANCE

Source	S.S.	d.f.	M.S.	Est. of*	F	Critical F
B	S_i = 15,723	3	4,574	$\sigma^2+9\sigma_i^2$	42.3***	$F_{.001,3,24}$ = 7.55
M	S_j = 15,854	2	6,927	$\sigma^2+12\sigma_j^2$	64.2***	$F_{.001,2,24}$ = 9.34
BxM	S_{ij} = 10,550	6	1,758	$\sigma^2+3\sigma_{ij}^2$	16.2***	$F_{.001,6,24}$ = 5.55
Resid.	$S_{a(ij)}$ = 2,582	24	108	σ^2		
Total	S = 40,709	35				

(6) Conclusions. The conclusions on the F ratios are:

1. The different brazing processes indicate a very strong effect on bonding quality.
2. The different materials also show a very strong effect.
3. The brazing processes do not effect the bonding in the same way with each material. Various combinations of the two factors produce very strong effects.

Further inferences can be drawn, such as: The true mean of each brazing process has 95 per cent confidence interval of the observed mean $\pm$ 2.065 ($\sqrt{108/3}$), that is, $\bar{x}_{i.} \pm 12.38$. We can ask ourselves which is the best combination (or combinations) of brazing process and materials, and is this significantly better? We shall not go into additional inferences other than to say that we can now attack many questions which were asked prior to the experiment.

Before leaving the topic of two factor experiments with more than one observation per cell, let us consider briefly a nested experiment and its model. In the previous example we noted that each level of one factor was crossed with each level of the other. For example, level 3 of Factor B (titanium) was always titanium for each of the four brazing materials. Suppose now we have an experiment where the second factor has no meaning except when considered within a particular level of the first. For example, consider an experiment involving 5 castings from each of which two samples were taken, and finally, two determinations were made. For the last column of Table 9, consider the number of samples which could have been taken as much larger than two. This is shown in Table 8 in which "Castings" is Factor A and "Samples" is Factor B. The distinction between Sample I and Sample II is meaningless except within a particular casting.

* Note change due to fixed levels.

Table 8 A NESTED EXPERIMENT

Samples \ Castings	1	2	3	4	5
I	x_{111}	x_{211}	x_{311}	x_{411}	x_{511}
	x_{112}	x_{212}	x_{312}	x_{412}	x_{512}
II	x_{121}	x_{221}	x_{321}	x_{421}	x_{521}
	x_{122}	x_{222}	x_{322}	x_{422}	x_{522}

The model for this example is

$$x_{ij\alpha} = \mu + e_{i.} + e_{j(i)} + e_{\alpha(ij)}$$

where $i = 1,2,\ldots,5$ $(= p)$

$j = 1,2$ $(= q)$ $N = pqn = 20$

$\alpha = 1, 2$ $(= n)$

The analysis of variance table for this example, using the same notation of Table 4 and Table 7, is given in Table 9.

Table 9 ANALYSIS OF VARIANCE TABLE FOR A NESTED EXPERIMENT

Source	S.S.	d.f.	M.S.	Expected Value
Castings C	$S_i = \frac{1}{qn}\sum_{i=1}^{p} T_{i+}^2 - \frac{1}{N} T_{++}^2$	p-1	$\frac{S_i}{(p-1)}$	$\sigma^2+n\sigma^2_{j(i)}+pn\sigma^2_j$
Samples within Castings S(C)	$S_{j(i)} = \frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{q} T_{ij}^2 - \frac{1}{nq}\sum_{i=1}^{p} T_{i+}^2$	p(q-1)	$\frac{S_{j(i)}}{p(q-1)}$	$\sigma^2+n\sigma^2_{j(i)}$
Residual	$S_{\alpha(ij)}$ = difference	pq(n-1)	$\frac{S_{\alpha(ij)}}{pq(n-1)}$	σ^2
Total	$S = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{\alpha=1}^{n} x_{ij\alpha}^2 - \frac{1}{N} T_{++}^2$	pqn-1 = N-1		$\begin{cases} p = 5 \\ q = 2 \\ n = 2 \end{cases}$

The last column of Table 9 indicates which F ratios to take. It tells us that "S(C)" mean square is tested against the residual. "S" mean square is tested against "S(C)" mean square if "S(C)" proved significant; otherwise, against the residual.*

* Again, we shall not enter into a discussion of pooling for variance estimates.

At this time it would be well to simply compare the last column in the analysis of variance tables which give us the variances being estimated. We consider crossed and nested, fixed and random (from a large number of levels). These expected values of the mean square estimates are given in Table 10. The levels of Factor A are i = 1, 2, ..., p and of Factor B are j = 1, 2, ..., q. There are n observations per cell.

Table 10 VARIANCES ESTIMATED BY M.S. IN A TWO FACTOR EXPERIMENT

I AND J CROSSED

Factor	Random i Random j	Random i Fixed j	Fixed i Random j	Fixed i Fixed j
A	$\sigma^2+n\sigma^2_{ij}+qn\sigma^2_i$	$\sigma^2+qn\sigma^2_i$	$\sigma^2+n\sigma^2_{ij}+qn\sigma^2_i$	$\sigma^2+qn\sigma^2_i$
B	$\sigma^2+n\sigma^2_{ij}+pn\sigma^2_j$	$\sigma^2+n\sigma^2_{ij}+pn\sigma^2_j$	$\sigma^2+pn\sigma^2_j$	$\sigma^2+pn\sigma^2_j$
AB	$\sigma^2+n\sigma^2_{ij}$	$\sigma^2+n\sigma^2_{ij}$	$\sigma^2+n\sigma^2_{ij}$	$\sigma^2+n\sigma^2_{ij}$
Residual	σ^2	σ^2	σ^2	σ^2

J NESTED IN I

Factor	Random or Fixed i Fixed j	Random or Fixed i Random j
A	$\sigma^2 + qn\sigma^2_i$	$\sigma^2 + n\sigma^2_{j(i)} + qn\sigma^2_i$
B(A)	$\sigma^2 + n\sigma^2_{j(i)}$	$\sigma^2 + n\sigma^2_{j(i)}$
Residual	σ^2	σ^2

The Latin Square. Consider another design called the Latin Square. This is a specialized design which has proven of value to industrial statisticians in the case of three factor experiments where the interaction can be assumed to be zero, and an equal number of levels of each factor is chosen. If we can assume the three factors independent we have a fraction of a larger experiment. This can still give us a great deal of information as an experiment in its own right or as a preliminary effort. We shall explain this by means of an example. *

The following coded data of Table 11 were the results of an experiment to compare five gasolines (A, B, C, D, E). The tests were conducted on each of five days (1, 2, 3, 4, 5). Five cars (a, b, c, d, e) were used. The arrangement of the design is such that the gasolines

* Chew - see reference No. 2.

and days were set up as rows and columns respectively. The third factor, cars, are randomly chosen to fill the cells with each car represented only once in each row and once in each column.

Table 11 A LATIN SQUARE EXPERIMENT

Gasoline \ Days	1	2	3	4	5	Totals
A	2(b)	0(e)	-2(a)	-3(c)	-1(d)	-4
B	8(d)	4(a)	3(e)	-1(b)	5(c)	19
C	1(e)	3(b)	5(c)	5(d)	7(a)	21
D	6(a)	1(c)	1(d)	2(e)	0(b)	10
E	7(c)	5(d)	2(b)	0(a)	4(e)	18
Total	24	15	9	3	15	64
Cars	a	b	c	d	e	
	15	6	15	18	10	

The analysis of these data proceed somewhat in the fashion of the main effects of the preceding models. The model of this design does not include interaction terms. This is an additive model of the main effects. The calculation of the sums of squares and analysis of variance table proceeds as follows: (The levels of each factor are equal)

For Days:
$$S_i = \frac{1}{p}\sum_{i=1}^{5} T_{i++}^2 - T_{+++}^2/p^2$$
$$= \frac{1}{5}(24^2 + 13^2 + \ldots + 15^2) - 64^2/25$$
$$= \underline{48.16}$$

For gasolines:
$$S_j = \frac{1}{p}\sum_{j=1}^{5} T_{+j+}^2 - T_{+++}^2/p^2$$
$$= \frac{1}{5}\left((-4)^2 + 19^2 + \ldots + 18^2\right) - 64^2/25$$
$$= \underline{84.56}$$

For cars:
$$S_k = \frac{1}{p}\sum_{k=1}^{5} T_{++k}^2 - T_{+++}^2/p^2$$
$$= \frac{1}{5}(15^2 + 6^2 + \ldots + 10^2) - 64^2/25$$
$$= \underline{18.16}$$

Total: $$S = \sum_{i=1}^{5}\sum_{j=1}^{5} x_{ij+}^2 - T_{+++}^2/p^2 \ *$$

$$= (2^2 + 8^2 + 1^2 + \ldots + 4^2) - 64^2/25$$

$$= \underline{214.16}$$

Residual: $$S_e = S - S_i - S_j - S_k$$

$$= \underline{63.28}$$

If we carry out the procedure as proposed earlier we have an analysis of variance table in which we test each of the main effects: gasolines, days, and cars. This analysis is given in Table 12.

Table 12 ANALYSIS OF VARIANCE TABLE

Source	S.S.	d.f.	M.S.	F-ratio	$F_{.05,4,12}$
Days	48.16	4	12.04	2.28	
Gasolines	84.56	4	21.14	4.01*	3.48
Cars	18.16	4	4.54	0.86	
Error	63.28	12	5.27		
Total	214.16	24			

As a result of the analysis we can say that there is some evidence of a mean difference due to gasolines. If there were particular pairings of grades beforehand, we could test these also. We might have said that we were not interested in differences among cars nor days. But by this experiment we separate out these possible effects and if there is a difference due to these, we may be able to establish it. Further, it is necessary to block these test runs within some units, such as days.

In conclusion, let us note that we considered a two factor experiment with more than one observation per cell. The models were presented and a detailed example carried out. Further inferences which can be obtained from our data were stated. Finally, we discussed briefly, by example, a particular fraction of a three factor experiment known as the Latin Square. This paper should be considered but a quick review

---

* We omit subscript k since by summing over i and j we have all the x values.

of some of the basic yet simple problems in experimental design. With an understanding of this paper many problems can be solved, but the field is still quite vast. The reader, if he makes extensive use of analysis of variance and experimental design, should feel it important to fortify himself with a few good books in the field. To just use a model problem or example as a cook book is sometimes dangerous. The successful experimenter goes further and investigates the philosophy and methodology behind these. There is no question that we have a powerful statistical tool. It must be used properly for maximum information.

References

1. Bennett and Franklin — STATISTICAL ANALYSIS IN CHEMISTRY AND THE CHEMICAL INDUSTRY (Wiley, 1954)

2. Chew — EXPERIMENTAL DESIGNS IN INDUSTRY (Wiley, 1958)

3. Cochran and Cox — EXPERIMENTAL DESIGNS (Wiley, 1957)

4. Cox, D. R. — PLANNING OF EXPERIMENTS (Wiley, 1958)

5. Davies — THE DESIGN AND ANALYSIS OF INDUSTRIAL EXPERIMENTS (Hafner, 1954)

6. Dixon and Massey — INTRODUCTION TO STATISTICAL ANALYSIS (McGraw Hill, 1957)

7. Hicks — FUNDAMENTALS OF ANALYSIS OF VARIANCE Part I, II, III. Industrial Quality Control Vol. 13, No. 2, 3, 4 (Aug., Sept., Oct. 1958

EXPERIMENTAL DESIGNS AND ANALYSIS OF VARIANCE

IV. SOME FACTORIAL DESIGNS *

Gayle W. McElrath
Industrial Engineering Division
Mechanical Engineering Department
University of Minnesota

INTRODUCTION.

It becomes apparent that the usefulness of the analysis of data depends greatly upon a systematic planning of the data in addition to the necessary assumptions of independence and distribution. This paper will introduce some of the ingredients of factorial designs together with a discussion of some of the reasons for confounding different effects and/or designing fractional replications. The mission of the paper is to present fundamental concepts which will be helpful as background thinking when designing experiments.

A classical accepted method of experimentation is to answer a single question about a single variable. There is an attempt to control all other "factors" except the one of interest. Many times the planning consists of trying first one idea, and if the one idea does not produce "satisfactory results" then try another idea. However, data based on this type of experimentation often consists of a number of disjoint pieces of information that cannot be easily put together. It is possible to design experiments such that all combinations of the different factor "levels" can be considered in order that their individual effects together with their interactions may be studied.

A GLOSSARY.

The SUBJECT of a designed experiment is the dependent measurable characteristic. Other variables which appear to affect the subject are called FACTORS. The factors of an experiment may be varied deliberately from trial to trial. The relation of subject to factors can be measured by proper planning of the given factors controlled at a given number of LEVELS. A TREATMENT is that combination of levels of all of the factors which are considered in a given trial. The numerical value corresponding to a given trial is the OBSERVATION or RESPONSE. When the levels of the factors have been chosen, and an experiment is designed to carry out one or more trials for each possible combination of the levels, the experiment is called a FACTORIAL experiment.

THE 2^3 FACTORIAL EXPERIMENT.

A factorial experiment containing k factors each at two levels, is often labeled as a 2^k factorial. In fact, when we take one observation per treatment there are 2^k treatments with $2^k - 1$ degrees of freedom to be parceled.

---

* This paper was presented at the
13th Annual Convention of the
American Society for Quality Control, 1959.

The 2^3 factorial designed experiment is one of the most instructive designs to give a quick realization of the concepts of main effects, interactions, and confounding. We will consider a 2^3 factorial design in which we wish to inquire about the effects of three independent variables (factors) A, B, and C upon the dependent variable (subject).

Consider the two levels of each factor to be given by A_1, A_2; B_1, B_2; and C_1, C_2. Figure 1 gives the 2^3 combinations for the treatments.

A factor	A_1				A_2			
B factor	B_1		B_2		B_1		B_2	
C factor	C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2
Treatment Combinations								

Figure 1. Display of the Data for a 2^3 Factorial Design.

The respective treatment combinations, from left to right, in Figure 1 are represented as

$$A_1B_1C_1,\ A_1B_1C_2,\ A_1B_2C_1,\ A_1B_2C_2,\ A_2B_1C_1,\ A_2B_1C_2,\ A_2B_2C_1,\ A_2B_2C_2.$$

If we take only one observation per treatment, we have eight observations. If we wish a better estimate of the experimental error we can repeat the experiment under similar conditions. The experiment has been REPLICATED.

Even though the symbols for the respective treatments, as given above, are adequate, the following is more convenient and conventional. Let the letters a, b, and c denote respectively the upper levels of the factors A, B, and C. (The previous notation was A_2, B_2, and C_2.) If the letter does not appear in the symbol for the treatment combination then it is understood that the factor not literally represented is contained at its lower level. For example, the treatment $A_1 B_2 C_2$ will be represented as bc in our new notation. We define the symbol (1) to denote the treatment where all factors are at their lower levels. In addition, now let A, B, C, AB, AC, BC, ABC denote the totals of the effects due to the factors and their respective interactions. For example, the total of the main effect due to factor A would be the difference between the sum of the values of the treatments containing the upper level of A and the sum of those containing the lower level of A. We write

$$A = abc + ab + ac + a - bc - b - c - (1)$$

and

$$B = abc + ab + bc + b - ac - a - c - (1)$$

$$C = abc + ac + bc + c - ab - a - b - (1)$$

A very interesting identity serves as a mnemonic device which may be used to determine the treatment combinations as follows:

$$A = (a - 1)(b + 1)(c + 1)$$

$$B = (a + 1)(b - 1)(c + 1)$$

$$C = (a + 1)(b + 1)(c - 1)$$

The rule for the plus sign (+) or minus sign (-) does not have to be labored at this stage but should appear to be quite obvious.

Also geometric devices may be used to aid in understanding that which otherwise might seem quite abstract. We may visualize the 2^3 treatments as the corners of a cube as described in Figure 2. The array of treatment

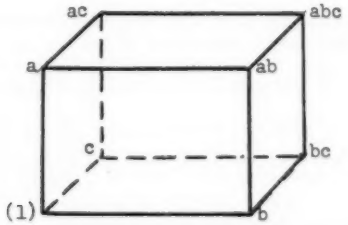


Figure 2. Geometric Representation of the Treatments of a 2^3 Factorial Design.

combinations of the 2^3 design of Figure 1 are represented by

(1), c, b, bc, a, ac, ab, abc

It seems quite natural to describe the total effects of the factors as previously defined when we identify the treatments with the artifice of geometric levels.

When the effect of one factor depends on the level of another factor, the two factors are said to interact and we have a TWO FACTOR interaction or, called by some, a FIRST ORDER interaction. To measure the total effect of the interaction of A and B, denoted by AB, we take the difference of the effect of A at B_2 and the effect of A at B_1. Therefore

$$AB = (abc - bc + ab - b) - (ac - c + a - (1))$$

$$= abc + ab + c + (1) - bc - ac - a - b$$

and

$$AC = (abc - bc + ac - c) - (ab - b + a - (1))$$

$$= abc + ac + b + (1) - bc - ab - a - c$$

$$BC = abc + bc + a + (1) - ab - ac - b - c$$

Again one has only to verify that the treatment combinations for the interactions are given by

$$AB = (a - 1)\ (b - 1)\ (c + 1)$$

$$AC = (a - 1)\ (b + 1)\ (c - 1)$$

$$BC = (a + 1)\ (b - 1)\ (c - 1)$$

To build a three factor interaction, ABC, consider the effect of level C upon the AB interaction and write

$$ABC = \left[(abc - bc) - (ac - c)\right] - \left[(ab - b) - (a - (1))\right]$$

$$= (abc + c + b + a) - (bc + ac + ab + (1))$$

$$= (a - b)\ (b - 1)\ (c - 1)$$

Thus we regard ABC as the difference between AB at the upper level of C and AB at the lower level of C. The three factor interaction is somewhat more complicated than the two factor interaction. In fact, higher order interactions become increasingly difficult to interpret in a real situation. Later we shall consider carrying out only a fraction of a larger factorial experiment and still obtain information about the main effects and lower factor interactions.

It must be quite obvious that extensions of the factorial design are possible. For example, a general case of 2^k factorial experiment is very possible. However, it does not take very much arithmetic to discover that each factor added to the experiment doubles the number of treatments. Thus for a 2^7 factorial design, 128 different treatments would be required and that by just adding an eighth factor increases the number of treatment combinations to 256. Modifications are in order which allow us to reduce the number of treatment combinations and at the same time retain the important estimates without determining the less important ones.

Another extension of the factorial design considers designs which involve factors other than at two levels each. The 2^k design may be

used to explore the relative importance of the k factors. This information may very well suggest further study of certain factors. Addition levels for these factors might be considered. Again there exists a formal analysis of variance of the data and discussion of methods to best estimate the experimental error. (10)

The advantages of the factorial design may be summarized as [6]:

(a) When there are no interactions the factorial design gives the maximum efficiency in the estimation of effects.

(b) When interactions exist, their nature being unknown, a factorial design is necessary to avoid misleading conclusions.

(c) In the factorial design the effect of a factor is estimated at several levels of the other factors, and the conclusions hold over a wide range of conditions.

CONFOUNDING.

In the previous section we mentioned the rapid growth of the number of treatment combinations as the number of factors k increased in a 2^k factorial experiment. A need for "confounding" certain effects arises when the number of combinations exceeds the capacity of the size of the blocks to be used for the factorial experiment. The important problem becomes how to split the treatment combinations among the blocks. Note that we are not reducing the number of treatment combinations, but are strategically assigning them to the blocks.

In order not to become involved in the detail of a higher order factorial experiment, consider again that we have a three factor experiment and each factor at two levels. Imagine that the limitations of the testing facilities force us to allocate the 2^3 treatment combinations to two blocks of four treatments each instead to one block of eight treatments. Let us assume that the decision has been made that the main effects together with the two factor interactions are of interest although the selection of the important factor effects is largely a technical matter. We can estimate these effects when we split the eight treatment combinations into two blocks of four treatments each if we sacrifice the three factor interaction, ABC. Previously we found that the total for the effect of the interaction ABC can be written

$$ABC = (abc + c + b + a) - (bc + ac + ab + (1))$$

Thus we may parcel the treatment combinations as follows in Figure 3.

Block 1	abc, a, b, c
Block 2	bc, ac, ab, (1)

Figure 3. Allocation of Treatments for a 2^3 Factorial to Confound the ABC Interaction with the Between Block Effect.

It should become clear that the effect of the three factor interaction ABC and the effect of a possible block difference cannot be separated. The effect of the three factor interaction is said to be CONFOUNDED with the between blocks effect.

If we would have decided to sacrifice a main effect, say A, in order to retain a measure of the three factor interaction, the allocation of treatments would be as in Figure 4.

Block 1	abc, ab, ac, a
Block 2	bc, b, c, (1)

Figure 4. Allocation of Treatments for a 2^3 Factorial to Confound the Main Effect A with Between Block Effect.

We mention that the concept of confounding can be extended to many types of designs, but the basic idea is well illustrated by the above 2^3 factor design in which the block size was restricted. Interest in this topic has produced much literature involving both mathematics and statistics. (2), (6), (10)

FRACTIONAL REPLICATION.

In some instances, it is quite impractical to run the 2^k treatment combinations necessary to give the complete information that a factorial experiment affords. It soon becomes evident that a large number of the $2^k - 1$ degrees of freedom are parceled out to the relatively high factor interactions. If the effect of these higher order interactions are likely to be small, it may be possible to arrive at an estimate of the residual variance which is sufficiently refined for our needs. However, when we reduce the number of treatment combinations we must be prepared to sacrifice some of the information that a complete factorial experiment affords. The consideration then becomes one of obtaining satisfactory estimates of the effects which are believed to be of interest from only a fractional number of the 2^k treatment combinations.

Actually, one may use any one of the blocks used in confounding a 2^k factorial design as a FRACTIONAL REPLICATE. However, this is rather a permissive approach to the problem. The order of logic involves deciding what information one might desire and then selecting the particular fractional replication that best gives the needed information. Frequently the main effects and the two factor interactions are of interest.

HALF REPLICATE OF 2^4 FACTORIAL DESIGN.

For purposes of illustrating some of the concepts of fractional replication consider a 2^4 factorial design. There are 32 treatment combinations. We wish to estimate the main effects and the two factor interactions. Without prolonged discussion, let us assume that the four factors are given by A, B, C, and D, and that we can find the total for the effect of factor A by expanding

$$A = (a - 1)(b + 1)(c + 1)(d + 1)$$

$$= \underline{abcd} + acd - bcd - \underline{cd} + acb + \underline{ac} - \underline{bc} - c$$

$$+ abd + \underline{ad} - \underline{bd} - d + \underline{ab} + a - b - \underline{(1)}$$

Also consider that the total effect of the three factor interaction BCD can be obtained by expanding

$$BCD = (a + 1)(b - 1)(c - 1)(d - 1)$$

$$= \underline{abcd} - acd + bcd - \underline{cd} - abc + \underline{ac} - \underline{bc} + c$$

$$- abd + \underline{ad} - \underline{bd} + d + \underline{ab} - a + b - \underline{(1)}$$

Now form a block of the eight common underlined treatments of A and BCD. We have designed a HALF REPLICATE of the 2^4 factorial design containing the following treatment combination:

$$abcd + ac + ad + ab - cd - bc - bd - (1)$$

The above half replicate estimates, as best it can, the main effect of A and does not distinguish this result from its estimate of the three factor interaction BCD. That is, the effects A and BCD are estimated by the same linear combination of treatments. The effect A is said to have an ALIAS which is the effect BCD.

Similarly the two factor interactions AC and BD can be estimated by the same common linear combination of treatments. We have

$$AC = (a - 1)(b + 1)(c - 1)(d + 1)$$

and

$$BD = (a + 1)(b - 1)(c + 1)(d - 1)$$

The half replicate is assigned the following treatment combinations.

$$abcd + bd + ac + (1) - cd - ab - bc - ad$$

and gives identical estimates of the effects of the two factor interactions AC and BD. Again, we say that AC has an alias BD. Of course, if the information from both blocks is obtained we have unbiased

estimates of the total effect of both AC and BD. It can be shown that there is a method of generating the aliases, and we find that our half replicate of the 2^4 factorial design has the following aliases which represent the seven degrees of freedom for the eight treatment combinations:

(A,BCD) (B,ACD) (C,ABD) (D,ABC)

(AB,CD) (AC,BD) (BC,AD).

From the above array of aliases we rightfully ask ourselves a question concerning the practicability of such a design. Unless we have information to the contrary, each two factor interaction has the possibility of being significantly large.

Another set of aliases possible from the half replicate of a 2^4 factorial design is as follows:

(A,BC) (B,AC) (C,AD) (D,ABCD)

(AD,BCD) (BD,ACD) (CD,ABC)

This set of aliases, in general, produces even less assurance than the former set. Here we have the main effects confounded with two factor interactions which also may be large. The advantages of fractional replication are better illustrated by 2^k factorial designs where $k \geq 6$. Even for $k = 6$, the higher factor interactions are relatively more likely to be negligible. For example, the main effects can have five factor interactions as aliases such as

(A,BCDEF) (B,ACDEF) ...

The two factor interactions have four factor interactions as aliases such as

(AB,CDEF) (AC,BDEF) ...

The three factor interactions have three factor interactions as their aliases such as

(ABC,DEF) (ABD,CEF) ...

The 31 degrees of freedom for the half replicate of the 2^6 factorial assignment are assigned as follows: 6 degrees of freedom to the main effects, 15 degrees of freedom to the two factor interactions, and 10 degrees of freedom to the 10 pairs of three factor interactions. The latter 10 degrees of freedom can be used to estimate the experimental error. If the fractional replication gives sufficient information, the experiment may be discontinued. If the results are not sufficient, it is still possible to complete the other half replicate for further information.

We have just considered the concept of a half replicate of a factorial design. In the area of fractional replications there has been much mathematical and statistical research to consider combinations of confounding and fractional replications of higher order, say one-quarter or one-eighthof the total replicate. (2), (6), (10)

SUMMARY.

Anyone who tries to make sense out of observational or experimental data is assuming the role of a statistician. However, the ability to scrutinize, to look beneath the surface, and to discern relationships from a given set of data is aided by scientific methods of planning, collecting, analyzing, and interpreting data. The modern methods of designed experiment and analysis deal with such questions as:

1. What factors are considered important and of interest?
2. What type of designed experiment should be used?
3. How shall a program of obtaining the data be planned?
4. How shall the data be analyzed?
5. What conclusions are we entitled to draw from the data?
6. How shall the experimental error be estimated and how reliable are the conclusions?

The problems associated with production and research experimentation are basically problems that demand an analysis of the variation of the data in an attempt to estimate certain effects and determine which variables are significant. By method of designed experiments, plans are available which consider the investigation of the effects of a number of independent factors at the same time. To this end, statistical methods are used to analyze the data from the designed experiment to calculate the total amount of variation which is sub-divided into categories traceable to definite sources. The relative sizes of these key categories of variation are compared with the size of the "experimental error."

The inferences drawn from designed experiments are valid inferences containing a measure of uncertainty. However, these inferences are rigorous when made by statistical analysis, since they include a probability measure of the amount of uncertainty involved. This achievement makes it possible to determine the reliability of the process of acquiring new knowledge by observation.

REFERENCES.

1. Anderson, R. L., and Bancroft, T. A. (1952), Statistical Theory in Research, McGraw-Hill Book Co., Inc., New York.

2. Bennett, C. A., and Franklin, N. L. (1954), Statistical Analysis in Chemistry and the Chemical Industry, John Wiley and Sons, Inc., New York.

3. Chew, V., (1958), Experimental Designs in Industry, John Wiley and Sons, Inc., New York.

4. Cochran, W. G., and Cox, G. M. (1957), Experimental Designs, John Wiley and Sons, Inc., New York.

5. Cox, D. R., (1958), Planning of Experiments, John Wiley and Sons, Inc., New York.

6. Davies, O. L., (Ed.) (1956), The Design and Analysis of Industrial Experiments, Hafner Publishing Co., New York.

7. Davies, O. L. (Ed.) (1957), Statistical Methods in Research and Production, Hafner Publishing Co., New York.

8. Dixon, W. J., and Massey, F. J., Jr. (1957), Introduction to Statistical Analysis, McGraw-Hill Book Co., New York.

9. Hald, A. (1952), Statistical Theory with Engineering Applications, John Wiley and Sons, Inc., New York.

10. Kempthorne, O. (1952), The Design and Analysis of Experiments, John Wiley and Sons, Inc., New York.

POLYNOMIAL CURVE FITTING ON ELECTRONIC COMPUTERS

by

Milton E. Terry
Bell Telephone Laboratories, Incorporated
Murray Hill, New Jersey

This paper discusses the problem of polynomial curve fitting by electronic computers. A solution involving Chebyshev polynomials is proposed which works well on the IBM 650 and 704, and is practical for desk calculators.

The fitting of a polynomial to a set of data falls into one of two cases. Either the observed values of the independent variable are equispaced or they are not. When these values are equispaced, it is well known that orthogonal polynomials exist in tabulated form or are available as recursion formulae. Standard procedures are available and straightforward except for some slight algebraic difficulty in recovering the coefficients of the equation in power series form. When, however, the observed values of the independent variable are not equispaced, and a k^{th} order polynomial is to be fitted, the construction of the k orthogonal polynomials corresponding to the n observed values is quite difficult in itself. When the number of points, n, is large, the procedure is expensive and involved even for large computers. When faced with the curve fitting problem with nonequispaced data, many experimentalists have tried to obtain the least squares estimates of the coefficients

of the power series directly, with varying degrees of success. The computing difficulty arises mainly from the condition of the square matrix A of the normal equation $A \cdot B = G$.

An objection may be raised immediately that in solution of linear equations and in the solution of these statistical problems of regression types it is not necessary to invert the matrix to reach a solution. Whether these problems are solved by getting the inverse of the matrix and then the solution desired or getting the solutions by essentially a step-wise reduction of the equations, if the data is such that the matrix is ill-conditioned, then the loss of accuracy will be progressive through the forward solution. Indeed, it is possible to find matrices so ill-conditioned that the diagonal elements in the forward solution actually become negative on limited accuracy devices. Thus, if a method of transforming the data can be devised which will allow a reasonable inverse, then it would appear that any of the other procedures for solving these problems would profit by the utilization of such transformations and conversely. Before discussing transformations to improve the inverse of a given matrix A, it is germane to raise the question of deciding when a matrix C is the inverse of a matrix A. Mathematically, the answer is trivial, $CA = I$. The answer is far from trivial when C has been computed and the product $CA = D$ has been calculated and the element d_{ij} are found to be

$\delta_{ij} - \varepsilon \leq d_{ij} \leq \delta_{ij}+\varepsilon$, where Kronecker's delta

$\delta_{ij} = 1, \; i = j$

$= 0, \; i \neq j$; and ε is a small positive number.

Is there any criterion by which one can stipulate the maximum acceptable value of ε in advance? Indeed, is there any other criterion for stipulating in advance the conditions for C being a satisfactory estimate of the inverse of A? As a start in exploring let us note that there are two distinct types of problems in which matrix inversion arises. The first problem may be characterized by the problems of multiple regression or the solution of n simultaneous linear equations in n unknowns, the second type of problem involves the exhibition of the inverse matrix as the end point, i.e., the solution. The first problem can be stated in matricial form as $A \cdot B = G$, where A and G are known matrices and B is an unknown matrix whose elements must be derived, the solution being written $B = A^{-1} \cdot G$. Consider C_0 as an approximation to the inverse of A, then $B_0 = C_0 \cdot G$ and $A \cdot B_0 = G_0$. Let us now calculate the two vectors with elements $g_{0_i} - g_i$, and $\frac{g_{0_i} - g_i}{g_i}$, when $g_i \neq 0$. M. B. Wilk[2] has conjectured that either one or both of these vectors may be useful in stipulating in advance a useful criteria for deciding when sufficient accuracy has been achieved. This criteria might be the specification of a value as an upper bound in absolute or percentage units which each

element of the appropriate check vector, $(g_{01}-g_1)$ of $(\frac{g_{0,1}-g_1}{g_1})$ must not exceed. Hence any matrix which yielded a check vector with elements $\hat{g}_1$ which satisfied the upper bound condition would be satisfactory. However, the use of $(g_1-\hat{g}_1)$ or $(\frac{g_1-\hat{g}_1}{g_1})$, $i = 1 \ldots, N$ as a test criterion for C_0 cannot be evaluated with statistical techniques. Even though g_1 may be a linear combination of the several values of a variate y, the $\hat{g}_1$ are a function of the computing machine and method of calculation as well as the data. Repeating the data on the same machine with the same method of calculation will yield identical $\hat{g}_1$. Since we know that the inverse of a symmetric matrix exists, we are evaluating the specific machine using a given method with respect to the calculated matrix. The use of the $\hat{g}_1$ does not, of course, preclude the computation of the AC_0 matrix as a check against gross errors. When the C_0 matrix is an unsatisfactory approximation to the true inverse matrix, and an optimal method of calculating the inverse has been used, the only way, in general, to improve the approximation without transforming the data is to resort to programming methods whereby much higher accuracy is brought into the calculations. This may imply double decision, triple precision, with perhaps the use of Hotelling's iteration rule.

Thus for problems of the first type, Wilk's criteria can be used to evaluate the over-all goodness of a matrix C_0 as an approximation to the inverse of the matrix A. For evaluating C_0 element by element no method is known

to the author, and hence for problems of the second type no answer is offered.

To illustrate the computing problem or the need for a transformation of the values of the independent variable in the nonequispaced case, we shall study the matrix whose elements are expected values arising as follows:

Let X be a sequence of values of x_α in the unit interval, and the values of x_α be uniformly distributed over the unit interval. Let a_{ij} be the element in the i^{th} row and j^{th} column of the matrix A and be equal to the expected value of x^{i+j-2}.

Then $a_{ij} = E(x^{i+j-2}) = \frac{1}{i+j-1}$. The determinant of A has been evaluated by Polya[1] as

$$\det |A| = \prod_{k=1}^{m} (\Gamma(K))^3 \Big/ \prod_{k=m+1}^{2m} \Gamma(k); \quad i,j = 1, \ldots, m,$$

which approaches zero rapidly. (For m = 5, $\det |A| \cong 1.6 \times 10^{-12}$.) If we consider the matrix A as a matrix of the normal equation $A \cdot B = G$, it is customary to reduce the order of the problem by correcting each element for its mean. Thus,

$$a^*_{ij} = Ex^{i+j-2} - Ex^{i-1}Ex^{j-1}$$

$$= \frac{(i-1)(j-1)}{i \cdot j} \cdot \frac{1}{i+j-1}; \quad i,j \geq 2$$

$$= 0 \quad ; \quad i,j = 1, \ i \neq j$$

$$= 1 \quad ; \quad i = j = 1.$$

If we now wish to solve for B, we must invert A. The abbreviated Doolittle is a standard method for inverting a symmetric well-scaled matrix. In this method at each major step of the forward solution all the row elements are divided by the lead element. The lead elements for the study matrix can be written as

$$A_{mm} = (\Gamma(m))^4/\Gamma(2m)\cdot\Gamma(2m-1),\ m \geq 1;$$

the first six values of A_{mm} being 1, $\frac{1}{12}$, $\frac{1}{180}$, $\frac{1}{2800}$, $\frac{1}{44,100}$, $\frac{1}{698,544}$. With only 8 digits of accuracy available it is clear that with such divisors, for $m > 5$ little or no accuracy will be left. Hence attempts to solve directly a polynomial least squares fit of order ≥ 4 on nonequispaced data using an 8-digit computer may be doomed. We need a transformation which for almost all sets of data will yield a matrix which can be inverted. We would like a transformation which yields equal precision in the estimates of all the coefficients and which is nearly an orthogonal transformation. The transformation defined in terms of the Chebyshev polynomials meets our criterion.

These polynomials can be defined as

$$T_j(x) = \cos j \arccos\ ;\ -1 \leq x \leq 1;\ \text{or}$$

$$T^*_j(x) = \cos j \arccos (2x-1),\ 0 \leq x \leq 1.$$

These polynomials have other properties of interest in curve fitting. The Max $T_j(x) = |\ \text{Min}\ T_j(x)\ | = 1$ within the interval (-1, 1) and these boundaries are attained for

all values of j, making these functions "equal ripple" functions. The Chebyshev polynomials have a higher average error in approximating a polynomial than the Legendre polynomials but have a lower maximum error. It has been shown by C. Lanczos[3] that if the tabular values of x_α are chosen at the zeros of the first neglected $T_n(x)$ polynomial, then the $T_k(x)$ polynomials are orthogonal with respect to summation for all $k < n$. If the data points are symmetric but not equispaced, this transformation will yield a matrix where $a_{ij} = 0$, $i+j$ odd.

We may use $T_j(x)$ as our transformation or we may select a value of k large enough to include most of the problems to be met, and try to find another transformation $H_j^{k*}(x)$ which can be used for all values of $k \leq k^*$.

The author has been unable to find a set $H_j^{k*}(x)$ which has a better conditioned matrix of expected values than the transformation set $H_j(x) = T_j(x)$. Since the power series expression for $T_j(x)$ has alternate zero coefficients when $-1 \leq x \leq 1$, it is better to apply the initial transformation to the independent variable x as

$$x = (2x'-x'_{max}-x'_{min})/(x'_{max}-x'_{min}).$$

The first nine values of $T_j(x)$ are given below, together with a suggested algebraic manipulation for programming.

j	$T_j(x)$	$T_j(x)$ (computer form)
0	1	1
1	x	x
2	$2x^2-1$	$2x^2-1$
3	$4x^3-3^x$	$(4x^2-3)x$
4	$8x^4-8x^2+1$	$8x^2(x^2-1)+1$
5	$16x^5-20x^3+5x$	$x+4x(x^2-1)(4x^2-1)$
6	$32x^6-48x^4+18x^2+1$	$(2x^2-1)\{16x^2(x^2-1)+1\}$
7	$64x^7-112x^5+56x^3-7x$	$x+8x(x^2-1)(2x^2-1)(4x^2-1)$
8	$128x^8-256x^6+160x^4-32x^2+1$	$32x^2(x^2-1)(2x^2-1)+1$
9	$256x^9-576x^7+432x^5-120x^3+9x$	$x+8x(x^2-1)(2x^2-1)(2x^2-1)(16x^4-12x^2+1)$

Whereas the study matrix $A = \{\frac{1}{1+j-1}\}$ will not invert correctly on an 8-digit floating decimal computer, the inverse of the matrix whose elements have been transformed inverts correctly in the sense that the product matrix of the matrix and its computed inverse is the identity matrix exactly when rounded to 8 decimal digits.

The remarkable gains in accuracy afforded by the use of this Chebyshev transformation as measured by the inversion of the matrix of expected values would seem to impute great value in the fitting of polynomials to non-equispaced data. Scrutiny of the diagonal elements of the inverse shows that on the "average" the coefficients of the transformed polynomials are measured with about the same accuracy. Further since the $T_j(x)$ are polynomials with integer coefficients, it is not only easy to derive the coefficients of the power series in x, but also little rounding error is involved. To obtain the power series in

the original x , considerable care should be taken in programming to control the rounding error. In general the use of the transformation $H_j(x) = T_j(x)$ can be recommended as excellent for fitting polynomials to nonequispaced data, and as acceptable otherwise.

ACKNOWLEDGEMENT

The writer would acknowledge and thank Miss S. J. Reed for the many computing programs involved in this study.

REFERENCES

1. Polya, G., Szego, G., Aufgaben und Lehrsatze, Dover Publications NY, 1945.
2. Wilk, M. B., Personal Communication, 1957.
3. Lanczos, C., Tables of Chebyshev Polynomials, Nat. Bu. Stds. Applied Math Series, 9, 1952.

SPECIFIC QUALITY PROBLEMS IN THE MANUFACTURE OF COMPONENTS FOR NUCLEAR POWER PLANTS

Charles M. Miller
Quality Control Manager, Bettis Plant
Westinghouse Electric Corporation

(U. S. Atomic Energy Commission Contract AT-11-1-GEN-14)

The Westinghouse Bettis Plant position with regard to the manufacture of nuclear components is unique in that the quality of products must be maintained by "remote control" since design and manufacture of equipment for the nuclear plants are procured almost entirely from suppliers outside our own plants. Remarks in this paper, as in the paper presented today by Mr. Greb, are primarily concerned with field inspection as carried on by customer inspectors working in the supplier's plant; depending on his facilities and to a large degree on his desire to produce quality products and upon his integrity in reporting those conditions which influence the ability of an inspector to judge compliance with specifications and drawings. While many specific problems are worthy of mention, this paper will be confined to three which by virtue of their nature can be used to illustrate special points relating to more general quality control problems.

The first of these problems pertains to the quality aspects of nondestructive testing and especially to the application, interpretation, and enforcement of radiographic inspection. It appears that the longer a method or process for inspection is used in a specific plant, the more lax is apt to become its enforcement and application to quality problems. Familiarity with the process leads technicians to take shortcuts and frequently accept deviations because they feel that their judgment is superior to the written requirements of the process specification. When a new inspection tool is procured at considerable capital expenditure, exceptional people are usually selected to take specialized training and acquaint themselves thoroughly with the process. However, these people move on to different jobs because they are exceptional individuals and the process is left to routine operators who never had the benefit of specialized training. These operators are considered to be highly skilled by virtue of several years experience but unfortunately this experience is often limited to repeated misapplications or misinterpretations. Twenty years experience in doing something wrong is much worse than one day's experience doing the same job correctly.

I believe that the general state of radiographic inspection in American industry today reflects a great many years of wrong experience and that radiographic technicians, inspectors, and supervisors need to review carefully the current output of their labs and evaluate the quality actually being achieved today.

The first step in the radiographic process is probably the most important. It is to decide on the technique which will produce a quality radiographic image. This step often gets little attention from qualified engineering or supervisory personnel. A part is delivered to the X-ray room and the technician decides on the spot how to make the picture. He does what he considers to be an adequate job based on the time allowed him and the facilities or equipment which happen to be

available. If the X-ray machine is available, he may use it. Otherwise, he may select one of the available isotopes. Frequently he gives no real thought to the matching of the proper equipment to the job and if his lab does not own the proper equipment, he will simply do the best he can with what he has. Only exceptional technicians understand the effects of using different types of radiation or techniques and apply their knowledge to the job at hand. The shipping foreman would not consider using a half-ton truck to move a three ton machine, but something similar to this is being done routinely in the radiographic lab nearly every day. Technicians are often hourly paid employes subject to union bumping or seniority displacement rules and turnover may be frequent. A man may have come into the X-ray Department just the previous month and about all he knows is that a source placed on one side of a weld and a film placed on the other produces a radiograph, but he is still given the responsibility of determining the technique and selecting the equipment to be used. The only acceptable solution to this problem is for competent personnel to prepare detailed technical instructions relating to equipment, films, identification markers, and the geometry of the part so that all of the variables are controlled. The technician then has a set of instructions to follow. The technician must have detailed information about the penetrameter which is to be used and its location on the part with respect to the film and source. The thickness of the penetrameter must be clearly defined or clear rules presented telling the technician how to determine the thickness of the penetrameter which he selects. All these things are very fundamental to the radiographic process and yet we find them uncontrolled far more frequently than otherwise.

The technician himself may not understand what constitutes a film of acceptable radiographic quality and often the man who takes the picture neither processes nor interprets it. Since he gives no information to the interpreter with regard to how he took the film, the relationship of the film to the part is unknown by the interpreter. However, interpreters universally interpret the film they get almost without regard to whether or not the film should be interpreted at all or actually represents the part in question. Most interpreters look at the penetrameter and will read the film if they can see the vaguest suspicion of its image. Rarely do we ever see the interpreter compare the film to a drawing in order to determine if the technician selected the proper penetrameter or located the film properly with respect to the section to be inspected. Almost never does the interpreter make even the simplest check to see whether he has films to cover the entire part or even counts them to see if all the station numbers are included. We have had films accepted which were so opaque that they could not be penetrated by the light from the most intense viewer we could find and others that were accepted at about the density of a piece of window glass. In each case the reviewer made the statement that the films were accepted because they showed no defects. This was true but it was also true that the parts had not really been inspected. One of the reasons for casual interpretation of films is that a reviewer may have to view as many as 500 to 600 films a day and keep up with administrative work in addition to the viewing. This heavy load makes it impossible for him to give each film an adequate review. Many reviewers scan the films only for obvious defects while the obvious things would easily be detected by relatively untrained people. Probably the most dangerous

defect in a weld or casting is a crack. Tight cracks may show up as only hairline indications or, due to their orientation, might escape detection completely. The reviewer must conduct his inspection in such a way that he will have the greatest opportunity to detect these faint crack indications.

Careful interpretation cannot be done in a casual way. A systematic method must be followed; first, evaluating the results of the technique; second, assuring that the radiography was actually taken in a way which would represent the part adequately on the film; and third, viewed with a light source which is capable of resolving minute indications on the particular films in question and only then giving the films careful enough scrutiny to enable indications of real defects to be observed.

While most of this discussion has centered on radiography, the same attention to the detail of the inspection process must be employed when doing magnetic particle inspection, surface penetrant inspection, eddy current, or ultrasonic testing. The simplest elements of the particular process must be kept under rigid control. The current used in magnaflux should be specified by competent personnel and meters and gages checked frequently enough to assure that the operator has a method of controlling his effort. We have even found people accepting parts being subjected to inspection when the inspection equipment was not even turned on. The operator was required to state only that he did not find defects. Penetrant inspections have been performed where a strong cleaning solvent instead of the recommended cleaner was used to remove the excess dye after its application to the part. Penetrant inspectors frequently do not understand the importance of allowing enough time for the dye to penetrate the defects. The danger of these kinds of errors is that the results obtained when the machine is not turned on, when the radiograph is as clear as a window pane, or the dye is not given time to soak into the crack are the same as if a good part was subjected to good inspection. Inspection methods which make an improper rejection of a good part are costly but not dangerous from a high quality integrity standpoint, but inspection methods which give negative results for a good part and equally negative results for a bad part must be under constant surveillance.

We deal primarily with two kinds of inspection agencies. First, the captive inspection group or quality department in a contractor's plant. The output of this department is responsible to management and depending upon the management attitude toward quality control and their surveillance, such a department can be regulated to obtain any desired degree of quality integrity. The status of training of personnel can be known by management and management can direct the staffing and control the facilities to attain the degree of quality desired. Frequently, managers do not make adequate use of knowledge at hand or do not take time to review it but at any rate, the responsibility for quality control is squarely on the company concerned. In order to give management the best opportunity to make proper decisions regarding product quality, quality control personnel must constantly feed realistic information and recommendations to them for consideration. A great many concerns (particularly smaller ones) depend upon professional testing or inspection labs for radiography, sonic testing, magnetic

particle, and penetrant inspection as well as analytical and physical testing of raw materials. These private labs have set themselves up as experts in particular fields. Their sales people represent their technicians as outstanding experts. Quite frequently these labs have exceptionally fine people in supervisory and management positions but the ability level of technicians performing the work often leaves something to be desired. When such agencies are employed, the quality responsibility then is lifted from plant management and becomes a subcontracted item administered through a purchasing department. Price, delivery, competitive buying practices, and even personal patronage quite frequently take precedence over pure quality concern. Labs will often resort to sloppy work or casual interpretation in order to avoid offending a customer or in order to meet hasty and unrealistic delivery dates. People purchasing such services frequently do not employ anyone who can adequately review the work being performed by the lab and have to depend blindly upon the lab for adequate inspection. I feel that the independent lab has a great responsibility to his customer to help him stay out of trouble and to forward first-rate realistic reports of conditions noted in products which are being inspected. Casual interpretations by these people can be extremely expensive if an item fails in the field or if a prime contractor or government inspector rejects equipment after it has been completed. Lest I be misunderstood, I recognize that there are a great number of competent labs operated by people of high integrity who will give meaningful and accurate interpretations when consulted on inspection problems. However, these high integrity labs find it difficult to compete with the others when the only consideration given to placement of orders is price and delivery. My plea to quality control personnel and plant managers is to be sure that your purchasing department selects competent vendors on a realistic quality basis and that you exercise caution when accepting the results of these purchases, checking and rechecking to be sure that the desired level of quality is being maintained. In summary, I can state that the enforcement of good inspection practices in the various nondestructive test requirements has been our most difficult problem. The critical requirements of this new nuclear industry have caused us to look extremely hard at materials and manufacturing processes in some segments of American industry. Frankly, we have not been very well satisfied with what we found. We are trying to stimulate, in the management of the companies with which we deal, an awareness of the importance of these processes. Constructive criticisms have been universally accepted and definite improvements have been achieved. We see evidence of this awareness spreading through other segments of industry and we are confident that this will continue to the benefit of our entire economy.

The second problem I wish to discuss involves visual and dimensional inspection of a special type of assembly operation. The nature of a nuclear power plant is such that maintenance and inspection of equipment after it is placed into service is extremely difficult. Parts are hermetically welded together with many welded seals making visual inspection difficult and the radioactivity present in an operating system makes accessibility available only to a limited degree. In the pressurized water type reactor there are high velocity flow conditions and very small channels so that loose parts carried along in the system can cause a great deal of trouble and could conceivably destroy the reactor.

Therefore, we specify that designs shall be made so that there are a minimum number of loose parts and such parts must be securely fastened or anchored to rigid parts of the plant so that they cannot get loose in the stream. The devices which may be used to lock pieces into place are definitely limited. We prohibit the use of devices which employ only friction such as lock washers, and we discourage the use of devices such as cotter pins which are severely distorted during installation to the extent that they might be broken. Locking methods such as prick punching which depend largely upon the skill of the workman and which are not easily inspected to see if they are adequate are also prohibited. The proper installation and assembly of such devices is a matter of primary concern for the integrity of the plant. Most of these kinds of things seem simple and perhaps not worthy of a great deal of attention, but simple things frequently receive the least attention and are apt to be overlooked entirely. After you have disassembled an entire plant to locate the one missing pin, 3/8" diameter and 2-in. long, and after you have had six $100,000 pumps severely damaged by a loose nut being pumped through the system, you are strongly inclined to set up rigid controls to see that such simple phases of assembly are accomplished properly. We have now required assembly data sheets which must be witnessed and initialed by assembly personnel and inspectors to certify that locking devices were installed and properly secured. In a complicated assembly which is not accessible for visual inspection after further assembly, we have had some suppliers photograph the assembly so that they would have definite proof of compliance with locking device specifications at the time of shipment. We have had pins which were designed to be press fit but which were so loose on assembly that they fell out when the unit was tipped. We have seen locking caps which were supposed to have been bent against the flat of a hex nut, but which were only bent against the corner and thus would not prohibit the nut from rotating, and we have had keys which slipped out of the keyways under no-load conditions.

I discuss this situation to illustrate the importance of infinite attention to detail in the nuclear program. Plants aboard ships are very closely confined so that the shielded compartments can be small and for this reason accessibility and maintenance is difficult. Land-based plants can be designed with more physical accessibility to the equipment, but land-based plants will probably be designed to permit higher levels of radioactivity and thus the accessibility for maintenance may be limited by the allowable radiation tolerance for maintenance personnel. This adds up to a need in the nuclear industry for extremely high standards of workmanship and materials. In many assembly operations, the integrity of the worker himself, and his ability and attention to detail, is the only real control on quality. When torque requirements are specified on nuts, only the man who held the torque wrench knows if the installation was properly made. Frequently, the work accomplished is immediately covered up, perhaps in the same operation as the one being performed, and only the assembler can tell if he did the job right. We have unusual requirements for cleanliness in the nuclear program and frequently must do work (welding, grinding, or polishing) on an already assembled unit. When such is the case, it is almost impossible to determine by inspection if dirt was allowed to enter the part. Yet the man who produced the dirt usually knows if he exercised enough control to keep the inaccessible portions of the unit

clean. In cases like this the quality control job is one primarily of worker education. We need to get the information about the requirements to the man doing the assembly and emphasize to him the reason for the strict requirements. Most persons will cooperate wholeheartedly with an inspector and quality requirement if they understand the need for it and have a feeling that the requirement is reasonable and meaningful. A machinist will usually exercise more care in producing two parts which he has to make fit together than he will if he makes the parts separately. This is because he understands what the parts have to accomplish and, armed with this knowledge, he exercises enough care to accomplish the desired result. I suggest that in any phase of assembly operations whether it be nuclear or missiles or otherwise, definite programs be established for acquainting assembly personnel with the real reasons why the tight requirements imposed are necessary. If you cannot justify these tight requirements, this is good evidence that the requirements are actually too tight and should be reviewed.

The third matter I wish to discuss pertains to the method of fixing responsibility for control of quality production. Since our experience in the nuclear power equipment program has been involved with government contracts, we as a prime contractor are influenced by the Department of Defense philosophy with regard to how we conduct our business. However, as a prime contractor we feel a strong sense of responsibility to work out the best possible methods of operation within the framework of general governmental policy. We have seen the time in this program when the customer inspected each item manufactured in its entirety whenever high quality or high integrity was an important requirement. In the earlier days of the nuclear program at Bettis Field, we attempted to supervise the manufacture of and perform the inspection on every piece of equipment manufactured for our program. When only one or two items were being manufactured on a tool room basis, it was possible to provide skilled manufacturing and quality personnel to perform this work and we felt somewhat delinquent if we were not able to have a manufacturing engineer, metallurgist, quality inspector and an expeditor present at every important stage of production. We witnessed qualification tests, performance tests, and confirmed dimensions by actual physical inspection. We reviewed the results of all nondestructive testing and witnessed the chemical and physical analyses to confirm that they were prepared under the right conditions. The first major construction effort of the nuclear propulsion program was conducted in the business atmosphere of a seller's market, intensified by the defense effort connected with the Korean conflict. We found it difficult to convince suppliers to participate in our program and many did so only after strong personal appeals by top management of our Company or by high ranking members of the Bureau of Ships. Many of the concerns which accepted our contracts did so on the basis of a patriotic desire to participate in a development program which had possibilities for significant contributions to our national defense. A good bit of the original work was contracted or subcontracted to small machine shops and other small producers who had neither design, manufacturing, nor quality control talent to meet the rigid requirements of our specifications. We used their facilities, supplementing them to a large extent by technical talent of our own. As the program grew, many of these small producers grew with the program and continued to expect the same technical assistance which they had received on first contracts. Not all of these producers were actually

small companies but smaller divisions or departments of large companies. These small segments of industry relied heavily on their customers for important decisions with regard to quality and, to a considerable extent, we encouraged them to do so since this gave us the most freedom to make changes and improve the equipment as fast as the development effort would provide new information. Many segments of American industry apparently operate in a similar way depending on the inspection or acceptability of tests by the customer to regulate the quality output of their plant. Old line, moderate size companies frequently grow static and fail to keep up with modern developments. We have dealt with foundries of considerable size which were pouring complicated alloy castings which did not even employ a metallurgist, chemist, or quality engineer. We have dealt with relatively large corporations who had never had a quality control organization of any sort and with departments of extremely large corporations who had never defined in writing to their people or customers what the quality control organization consisted of or who was responsible for it. We obtained a quality control organization chart from one concern that listed as Chief Inspector a man who had been dead for two years, and in another plant we found a disagreement between two quality inspectors working side by side as to who their boss was.

As long as customers were willing to supervise the quality effort themselves and accept the quality of products delivered, management was reluctant to improve conditions since this meant spending considerable money and many of these same companies were in an unfavorable profit position on nuclear work due to a number of reasons. However, a prime contractor usually cannot have his field inspector adequately trained in every phase of quality control and it is impractical to have a specialist visit a plant to conduct relatively short inspections. Therefore, we depended more and more as the volume of work increased upon one relatively inexperienced customer's field inspector to maintain the quality effort in an entire plant to the same degree which had been prevalent when only one segment of that plant had participated in a highly supervised effort. Likewise, the supplier depended to a considerable extent on the customer's representative in the plant to control quality and insure adherence to his bid specification. Since he had no organization of his own to do this, he accepted pretty much what the customer inspector had to say. We finally faced up to the problem with a knowledge that it was impossible for a limited number of customer inspectors to inspect everything in a large number of manufacturing plants and we either had to greatly reduce the extent and degree of inspection on a per-piece basis or else attempt to hire increasingly large numbers of field inspectors. The latter seemed impractical and so we have initiated a major effort to educate suppliers with regard to their responsibilities for quality control and to recommend specific changes and improvements within their organizations to accomplish the desired level of quality enforcement. Beginning about a year ago, the Bureau of Ships representatives involved in the program of the Navy nuclear power plants, with the cooperation of the several prime contractors involved in Navy nuclear propulsion work, started a series of quality evaluation visits to subcontractors' plants to review their quality control organization procedures and activities in a critical way, making constructive recommendations wherever observations indicated improvements could be made. These visits have been highly successful

from the prime contractor's point of view and have enabled us to establish contact with high level management in suppliers' plants concerning quality matters. Such contacts had not been easy to attain until there was a concerted effort on the part of more than one customer to impress plant management with their desire for him to establish a feeling of real quality responsibility within his plant.

When dealing with the highly complex products involved in our defense industry today, a manufacturing concern needs to establish a well defined system which is capable of determining the product quality and then must keep accurate records pertaining to results of inspection so that he is in a position to prove compliance with complicated specifications at the time equipment is shipped. The ultimate goal in the Department of Defense quality control philosophy, as we understand it, is to purchase along with products the output of such a quality control system and to inspect the results of the quality control system itself rather than to perform detailed inspection of the product. As a prime contractor we have set a goal of limiting in-process inspection to those portions of the product which cannot be adequately inspected at the time of shipment and then to conduct final inspection of the product and inspection records, demanding evidence and proof of compliance with applicable specifications at the time equipment is ready for shipment. When the customer's acceptance is withheld until the product is completed, the supplier must satisfy himself of the quality at the various in-process stages or else risk the costly investment of additional manufacturing expense in parts which will ultimately be rejected by the customer. We feel that sometimes a supplier offered defective material to an inspector at an in-process stage knowing that if it somehow got past the inspector at that point he would be under relatively small risk to complete and ship the item. On the other hand, if the customer's inspector rejected it at that stage, the supplier would simply repair or replace the part and would have had no additional expense as a result of his offering it to his customer. When inspection is limited to final-type, the supplier will be under a constant threat of rejection of valuable equipment unless he really takes every step to assure himself of product integrity at each and every stage.

A little earlier I mentioned the fact that many suppliers for the nuclear program were in an unprofitable situation and this fact has contributed to our quality problem perhaps more than any other single thing. Rarely have the suppliers of nuclear power plant equipment established separate departments or divisions in which to handle the nuclear work and it has been mixed in with standard commercial production to a far greater extent than seems to be the case in the aircraft, ordnance, and missile fields. Many suppliers have not geared themselves to read and understand the specifications let alone establish facilities to comply with them. Because of their failures to read and understand these difficult specifications, many companies bid unrealistically low. Subsequently, because of the considerable inspection effort by prime contractors, vendors were ultimately forced to comply with the details of the specification and it was impossible for them to make the product within the bid price. A number of suppliers were intrigued by the publicity which attended the birth of this new industry and decided to buy into the business by quoting prices which were known to be under costs in order to gain experience and become qualified as a

supplier to the program. Even having read the specifications, many suppliers were so unfamiliar with the limitations of their own capabilities that they improperly took contracts for the highly specialized products involved. Because of one or a combination of these reasons, a good number of bidders supplying equipment to the nuclear power program lost money and this fact had a great influence on the problems of quality enforcement. Profit conscious managers were forced to invest in additional inspection hours or inspection equipment for products which were being made at a loss. Profit conscious foremen and production personnel tended to delay the scheduling of unprofitable equipment which led to increasingly late deliveries and consequently greater intensification of production effort. Plant management was convinced by high level expediting that they had to make delivery at any cost. Under such conditions there was a tendency to try to justify the acceptance of borderline or slightly inferior apparatus and the quality conscious individuals in the nuclear program have had to be constantly vigilant to maintain the established standards. I am sure the only firm foundation for the purchase of high integrity, high uniform quality complex equipment is a healthy profit situation on the part of the manufacturer. Purchasing agents would do well to consider this as a major factor in placing such orders.

By sheer hard work and constant attention to nearly every operation, we have been able to maintain the highest level of quality and I consider the standards prevalent among suppliers for the nuclear program to be among the highest in industry. I have evidence of many cases where the establishment of high standard quality programs has led to overall improvement in the quality level of products, both commercial and military, coming out of a supplier's plant. In some cases this increased quality level was attained simply through better management and better education and was done at no increase in cost and conceivably is attendant with reduced costs. We have certainly seen giant strides in the enforcement of quality requirements in our industry by producers. In some cases, we have accomplished improvements in material quality which basic supoliers said could not be done and by development of their production processes they were able to give us the quality we wanted and all of their customers have benefited. We are proud of the quality contributions made by the nuclear power industry. Though only a few years old, its impact across a large segment of production capacity has been significant. We hope that our efforts have made the quality enforcement jobs of each one of you easier and have given you added incentive and ability to sell high integrity quality programs to your management.

EVALUATING INSPECTION COSTS

Paul E. Allen
Manager Quality Control & Customer Service
Beech Aircraft Corporation

Evaluating inspection costs is a project in which we are all engaged at all times. We may be doing it consciously or unconsciously, systematically or haphazardly. To many of us it is not an inspiring phase of our responsibility. However, there is no question about its being one of the most important quality control management functions.

A check of the AIA Quality Control System Study for 1958 shows that 83 out of 91 participating companies have a budget program under which the Quality Control Division operates. The same survey shows that 86 out of the 91 companies use sampling plans at various stages of their inspection program, and many of the other well-recognized quality control cost control and cost reduction programs are being used extensively. Therefore, in this talk, I will deal with some of the specific areas where we have made inspection cost evaluations with worthwhile results.

PREDICTING INSPECTION COSTS:

A common problem related to evaluating inspection costs is predicting costs for continuing projects, and particularly for new projects. This is usually done by use of past history, with adjustment for differences in work mix, production rate, etc.

WORK MIX VERSUS INSPECTION COSTS:

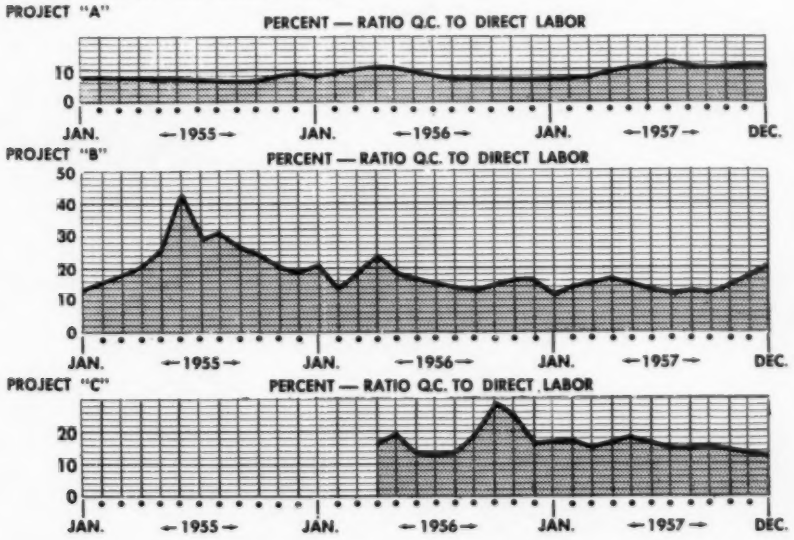


Figure 1

The effect of work mix on inspection costs is illustrated by Figure 1. This shows the quality control labor requirements on several different projects. These labor requirements are expressed as a percent of direct labor serviced--the work was of a similar nature, but on different projects for different prime manufacturers.

These different rates of quality control labor used by the same manufacturer on production for three different prime contractors indicate that work mix can be a very important factor in cost predictions.

PRODUCTION PHASE VERSUS INSPECTION COST:

The production phase also has a decided influence in predicting quality control costs. The learning curve for quality control shown in the following charts represents a definite new product pattern, any major change in ratio of new products to regular products would affect the quality control costs.

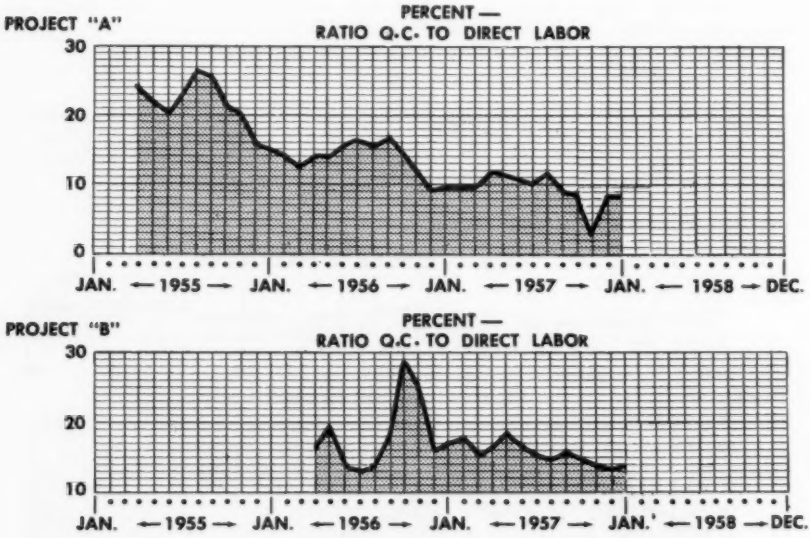


Figure 2

For example, assume that 100% of one's workload was on new projects during the first 12 months of production, with a control manpower requirement similar to "A", above-the-average control cost would be 21% of direct labor serviced; however, by assuming the other extreme that 100% of one's workload was on production of several years' run, such as the above Project "A" during 1957, the average control cost would be 9% of direct labor serviced.

PRODUCT FLOW RATE VERSUS INSPECTION COST:

Production flow rates affect the quality control costs, as indicated by the following chart which reflects increased costs with reduced production rates:

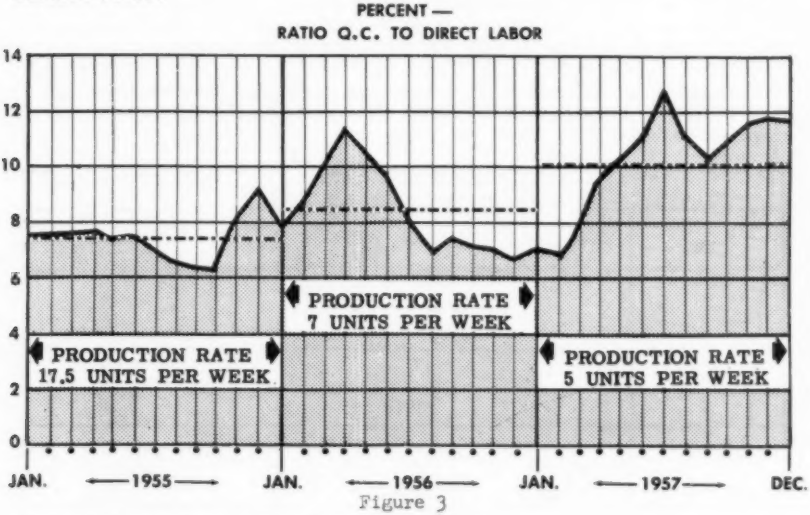


Figure 3

PRODUCTION PHASE-OUT VERSUS INSPECTION COSTS:

The effect of work phase-out can be of equal importance on costs along with work mix and new production, as indicated by the following charts on two projects that phased out:

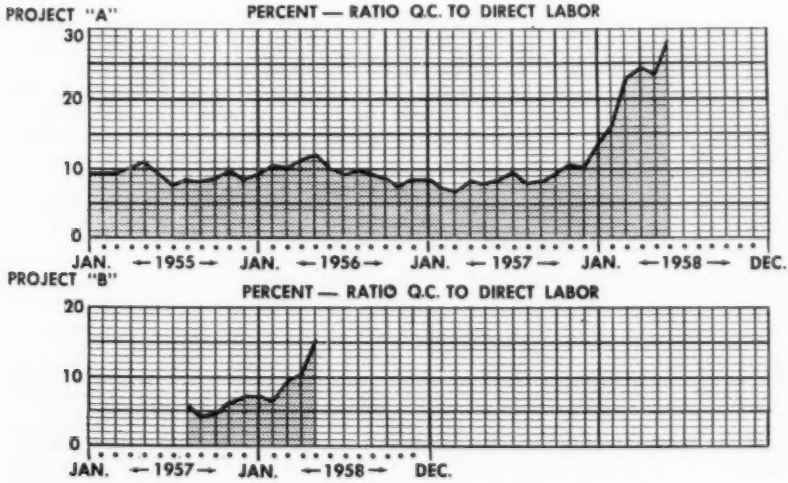


Figure 4

EFFECT OF "LOT SIZE" ON INSPECTION COSTS:

In the airframe industry, most receiving inspection, fabrication inspection, processing inspection and small subassembly inspection is performed on lots, as against continuous flow inspection. The inspection cost per unit usually proves to be influenced to a great extent by the lot size--i.e.: the larger the lot, the lower the unit cost.

Figure 5 shows changes in inspection costs per unit when the lot size changes. This example is taken from receiving inspection where sampling is used.

INSPECTION COSTS
AS INFLUENCED BY LOT SIZE

ITEM	PERIOD	AVERAGE LOT SIZE	UNIT INSPECTION COST IN CENTS
A	1	11,312	.017
	2	9,358	.021
	3	6,499	.030
B	1	87	2.000
	2	66	2.636
	3	35	4.971
C	1	147	1.299
	2	34	5.617
	3	16	11.937
D	1	23	9.060
	2	9	33.222
	3	10	29.900

Figure 5

There are, of course, many other factors that must be considered when a company makes a determination of optimum lot size. However, we find that it is common for those people who make this determination to ignore entirely, or underestimate, the increasing inspection cost factor (and other similar indirect expense) when small lots are processed at frequent intervals. We do not propose that inspection or quality contro should control lot size, as this is not a quality control function.

It is important that you make sure that the inspection cost relationship to lot size is understood and considered by those people or departments who determine optimum lot size.

IN-PROCESS INSPECTION VERSUS END-ITEM INSPECTION:

This is an inspection cost area that must be studied carefully by each company, and on each project. There is no one answer that is optimum for all cases. However, this only emphasizes the need for individual studies. After considerable study, we have concluded that as a policy in our plant, a high degree of "in-process inspection" by the inspection department is not a good investment. Certainly, in-process inspection is necessary and desirable, so perhaps an example of what is meant by a high degree of in-process inspection would help clarify the

point. Let us assume we have a machined part--the shop routing could be as in "A", with "in-process inspection", or as in "B", with end-item inspection:

Figure 6

At first analysis, it might appear that Example "A" would pay off in early defect detection, thus saving machine hours, etc. We have found that in most cases this is not a significant saving, and over a period of time, it usually has the undesirable result of having manufacturing dependent upon inspection for controls which rightly are manufacturing supervision responsibilities.

This same analysis must be made of assembly line operations, and here you will find the largest potential area for savings, but you will also find it a more difficult problem. Most manufacturing people will, at first, prefer that inspection "buy off" each job, operation, or installation as soon as it is completed and signed off by manufacturing. This can be done, but from the inspection cost standpoint, it certainly is not the economical way to do it. Much more efficient use of inspection manpower can be realized if inspection check points are designed

which will allow the inspector to check a group of items, a complete system, or a complete installation at one time.

Again, it appears that the item-by-item buy-off would have the advantage of early detection; the workman who did the work can be corrected immediately, before he repeats his error, etc. Our studies reveal that in most instances, the apparent gains are again more than cancelled out by the undesirable result of having manufacturing dependent upon inspection for functions which rightly are manufacturing supervision responsibilities.

I hasten to point out that the "in-process inspections" discussed here do not include defect prevention techniques such as statistical control charts, first piece inspection, etc.

EFFECT OF TYPE OF ENGINEERING REQUIREMENTS ON QUALITY CONTROL COSTS:

In examining areas of inspection costs, we found that about 25% of our costs were being expended in fabrication inspection. A check against the AIA Quality Control Systems Study reveals the average of the 31 airframe companies reporting were expending about the same percentage of their effort in these areas.

A careful analysis of the inspection being performed revealed that a great deal of time was being expended in checking myriads of detail characteristics to standard tolerances (usually standard "title block" tolerances). An analysis of the need for complete conformity control over parts to assure that they conformed to all engineering requirements led us to the realization that a very large percentage of engineering drawing dimensions are primarily for the purpose of aiding in tool and jig design and inter-group coordination, and do not contribute to the end products' quality, serviceability, interchangeability or reliability.

With this thought in mind, we began negotiating with Engineering to classify all drawing characteristics into Critical, Major, Minor "A", and Minor "B" Classifications.

At this point I should make it clear that the term "Classification of Characteristics" is not the same as "Classification of Defects". In "Classification of Defects" the tolerance on the characteristic is unchanged, and a lower classification of the defect merely means that a higher percent of defectives is acceptable. In "Classification of Characteristics" the engineer is actually changing the tolerance on the characteristic.

As a result of preliminary experiments in this line, we found that in one group of machined parts, Engineering considered that 43% of the dimensions could be ignored as long as the part made a satisfactory assembly, and on a group of sheet metal parts, 62% of the dimensions could be ignored as long as the part made up into a satisfactory assembly.

It was apparent from this preliminary analysis that such a program would result in some sizeable man-hour savings in inspection.

At Beech we classify characteristics as follows:

ENGINEERING CLASSIFICATION	ENGINEERING ACTION	INSPECTION ACTION
CRITICAL Those characteristics which, when exceeded, vitally affect safety.	No change in tolerance.	Normal Inspection.
MAJOR Those characteristics other than Critical which, when exceeded, are important to the function, interchangeability, utility, service life, appearance, etc., of the end product.	No change in tolerance.	Normal Inspection.
MINOR "A" Those characteristics, other than Critical or Major, whose standard tolerances can be exceeded by a specified amount as long as the part makes up into a satisfactory assembly.	Tolerance changed by a specified amount on the drawing. Example: ± means tolerance doubled both under and over nominal as long as the part makes up into a satisfactory assembly.	Normal Inspection using the expanded tolerance.
MINOR "B" Those characteristics, other than Critical, Major or Minor "A" which require only general adherence to specifications to produce a satisfactory end product.	Characteristic specified Minor "B": Dimension tolerance can be ignored as long as the part makes up into a satisfactory assembly.	Characteristic is not inspected.

Figure 7

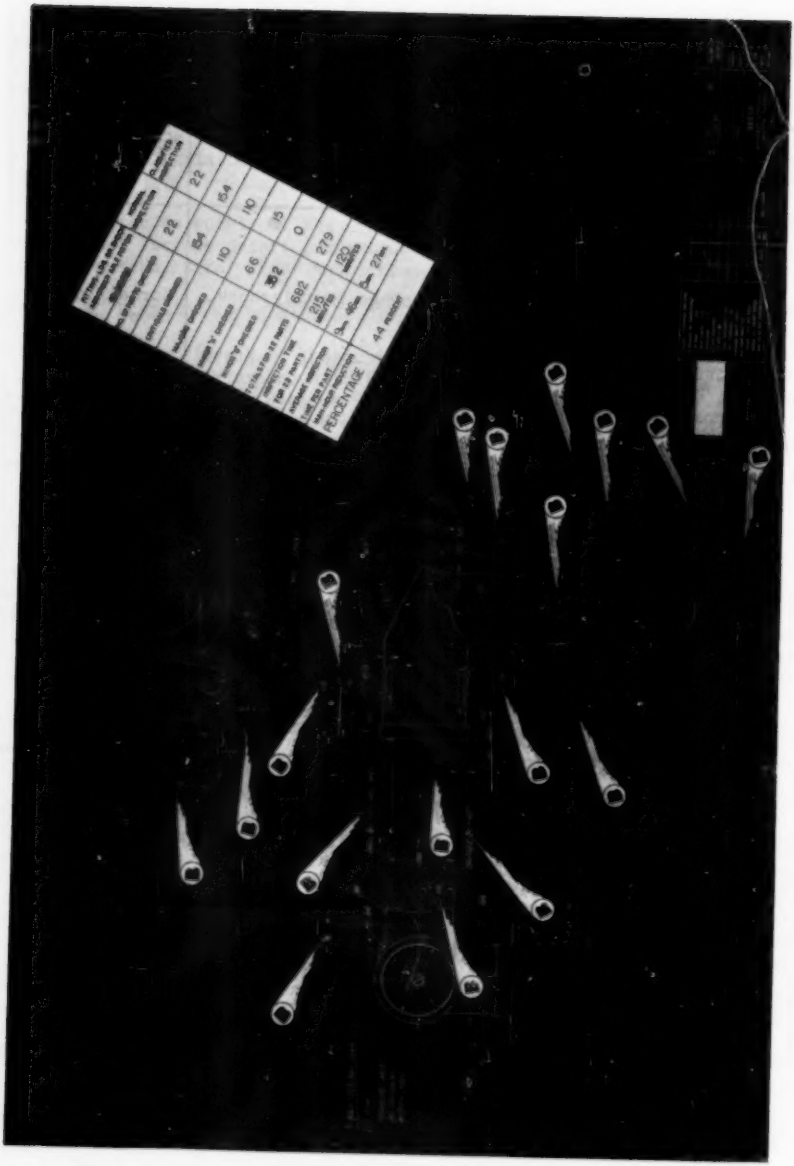


Figure 8

An example of the basic engineering classification as applied to a specific part is shown in Figure 8 of this report. A machined part is shown which reflects that 18 dimensions, or more than 50% of the total part dimensions, do not need to be controlled as long as the parts make up into a satisfactory assembly. A time study on one order of 22 of these parts indicated that it would require 215 minutes for one inspector to check all characteristics; however, he could check those classified as needing control by Engineering in 120 minutes--or there was a man-hour savings of 44% realized through use of the classified method.

The savings in inspection time paid off the engineering time spent in classifying the drawings in a very few cycles of parts. Inspection savings on subsequent cycles were pure profit. In this project, we found that the by-products were more valuable than our original goal. The by-products were:

1) Savings to the shops on reworking characteristics that were acceptable after classification.

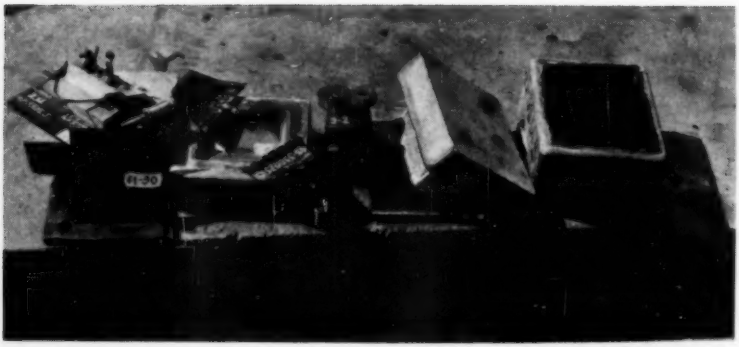
2) Savings to Tooling on tool rework and tool development to hold a characteristic to close tolerance when classification opened up tolerances.

3) Savings in original tooling costs on items that had characteristics classified, by permitting the design of less precise and less costly tools.

4) Savings to Quality Control and Engineering in processing rejections that are made unnecessary by the classification of characteristics.

An example of over-all savings to be realized from the "Classification of Characteristics" is shown with the aid of Figure 9. In this case, there were 95 pieces in a sequence of parts formed from the nine dies pictured. If these 95 pieces were processed through 100% inspection to all drawing dimensions, there would be ten rejection tags written at a cost of about $150, and if the tools were to be reworked to fabricate the parts to meet all basic drawing requirements, the rework cost would be about $1,200. In this case, exact conforming parts would be of no better quality than the ones already produced; therefore, any drawing changes, salvage expenses or tool rework would be a waste of time, money and effort. Here again, the Engineering time required to classify the characteristics is insignificant when compared to the savings realized.

The over-all savings to the company to be realized from "Classification of Characteristics" are so impressive that the Engineering Department has now "joined us" and are enthusiastic about the program. All new projects now have characteristics classified on the original design, and all old projects have been screened and classified by Engineering.



SERIES SAVINGS IN TOOL REWORK - WINDSHIELD DEFOGGER ASSEMBLY TOOLING

WINDSHIELD DEFOGGER ASSEMBLY PARTS

Figure 9

In a recent research project evaluating the effectiveness of this program in controlling critical and major characteristics, a lot of parts with known defects was checked by 17 inspectors from several companies where each man inspected the same lot of parts to 100% drawing characteristics and to the classified characteristics. An average error rate of 44% on Major and Critical Characteristics was found when all drawing characteristics were checked, and an average error rate of 17% on Major and Critical Characteristics when only the classified characteristics were checked. This was a staged test with some very difficult tricks used in putting defects in the coded parts; however, the results indicate an improvement in catching Critical and Major defects when the number of characteristics to be checked is reduced.

Therefore, we say that the type of engineering or drafting practice has a very significant effect on quality control costs, and should be given careful attention in any quality control cost analysis.

CONCLUSION:

Obviously these are only a few of the areas that should be evaluated in any program for evaluating inspection costs. It is one of those undesirable types of program that never ends. The check points as to whether or not you are doing a good job are not easily found. We have found the AIA Quality Control Survey to be a good general guide to tell us how we are doing with respect to others engaged in similar business.

So it boils down to this: Evaluating Inspection Costs is like shaving--you may do a good job of it today--but, you know it will have to be done again tomorrow!

BENEFITS OF TIME RECORDING FOR PRODUCER AND CONSUMER

W. K. Warner
Chief, Development Evaluation Section
Missile Division, North American Aviation, Inc.

Operating time records have been employed for many years in the aircraft industry. During the past few years, and especially since the growing complexity of equipment has focused attention on the problem of reliability, there has been an increased emphasis on both the recording and the statistical treatment of time data.

A certain degree of unreliability exists in all aircraft and missiles. This is apparent in the extensive maintenance facilities required, in the percentage of weapons which can be maintained in an operational readiness condition, and in the number of missions aborted because of equipment failures.

Time records, and the time per failure data derived from them, are one of the most powerful tools available to us in our effort to improve equipment reliability. A great deal of progress has been made in the use of this tool, and the benefits to the producer and consumer in development and production economy, and in the earlier availability of more effective weapon systems, are literally incalculable.

In the military equipment field no distinction can be made between producer and consumer benefits. Production economy does not mean a larger profit for the producer, the saving is always passed along to the consumer. The efficient producer hopes he will be awarded new contracts, but if he is, it is because the consumer thinks the producer's efficiency will be a consumer benefit. The more rapid development of more efficient weapon systems is not really a benefit to the consumer; it is the producer who ultimately benefits when the weapons are used to protect him, his home and factory.

Thus, in the military equipment producer-consumer relationship, what benefits one, always benefits both.

With the advent of highly complex modern aircraft and missiles containing perhaps a hundred parts where only one stood before, it quickly became apparent to producer and consumer alike, that new tools, new methods, new standards for design, production and quality control, were required to develop and produce equipment which could be both very complex and highly reliable.

Many thousands of words have been written defining the "reliability problem." The following simple statement seems to encompass all of them.

Excluding human errors which may occur in maintenance or operation of equipment, there are two, and only two, factors which make equipment reliable or unreliable. One is the design safety margin against <u>all</u> kinds of stresses. The other is the degree of control exercised over <u>all</u> production variables in workmanship and materials.

As shown in Figure 1, these factors are interdependent. Large safety margins are required if control of production variables is not good. Very good control of variables is required if safety margins are not high.

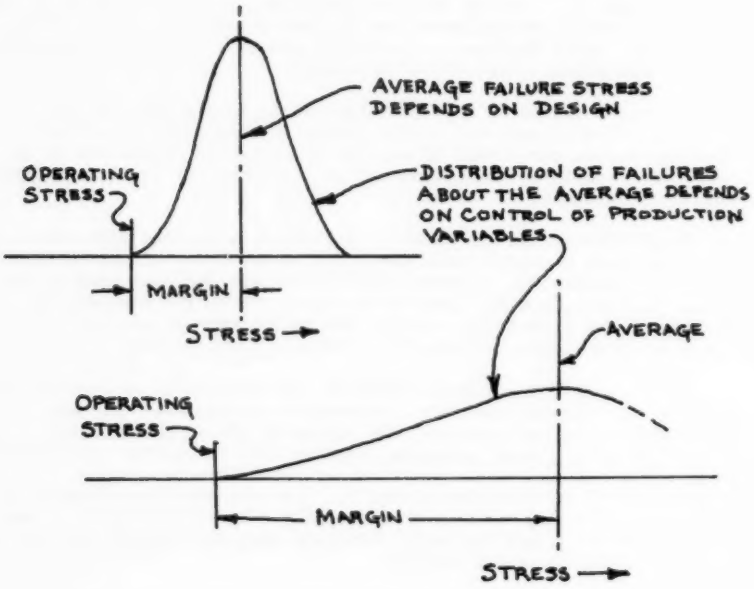


FIG. 1
DESIGN SAFETY MARGIN VS. CONTROL OF PRODUCTION VARIABLES.

While these statements apply for any kind of stress, environmental, handling, storage, etc., nowhere is the truth more apparent than when operating time is taken as the stress.

If a number of supposedly identical units from production are operated simultaneously, (either continuously or intermittently) until they fail, they will usually turn out to be not identical at all as far as time to failure is concerned.

A good correlation can be observed between the complexity of the equipment and the spread in the distribution of time to failure. When the equipment is more complex, a larger number of workmanship and material variables exist. Since these variables cannot be completely controlled, it follows that complex units are not produced as nearly

identical to each other as simple parts, and will display a wider spread in their time to failure distribution.

An immediate consequence is that engineering is forced to design complex equipment for very long average life in order to accommodate the wide distribution about the average and still not encounter too many failures at the time designated "operational stress" in Figure 1.

The probability of survival of aircraft components for a given period of time is sometimes expressed as the exponential function $P_s = e^{\frac{-t}{m}}$, where t is the time of interest and m is the mean time to failure for the equipment, prior to wear-out. The need for a long average life for high probability of surviving even short periods of time is shown in Figure 2, which gives the exponential probability of survival of a given component for the relatively short period of 1/2 hour, when the mean time to failure is assumed at various values.

Note that for a .9995 probability of successful operation during any given $\frac{1}{2}$ hour of operation, which is a common reliability requirement at the aircraft or missile component level, the mean time to failure required is about 1000 hours.

$P_s = e^{\frac{-t}{m}}$

P_s	.96	.99	.995	.9995
m	10 HRS.	100 HRS.	500 HRS.	1000 HRS.

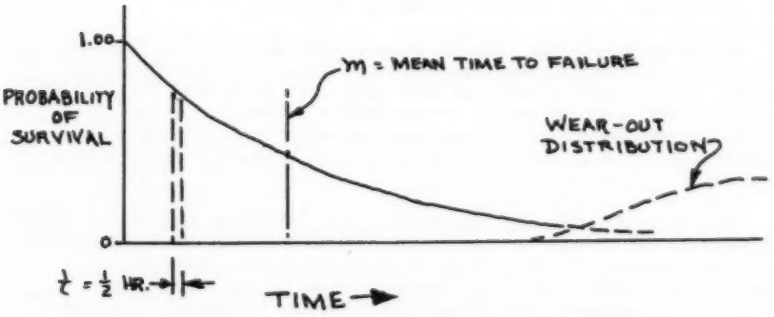


FIG. 2

PROBABILITY OF SURVIVING ANY ½ HOUR PERIOD AS A FUNCTION OF MEAN TIME TO FAILURE

Time records on systems and components can help determine what degree of control is being exercised over production variables. The plot of average time per failure, and the distribution of failures in time about this average, not only measures the effectiveness of control over production variables for each component, and pinpoints the areas where controls are most ineffective; it also tells the engineer what design margins he must provide to be commensurate with the state of the art of controlling production variables in the type of component being designed.

In airframe structural design, theory and experience have established standard safety margins with respect to structural stresses which make adequate provision for production variations. At the same time the controls over structural workmanship and material variables have been well developed so that successive production components are not too different from each other. Thus, although strength does vary somewhat from one airplane wing to the next as they are produced, even the weakest wing produced is still more than able to carry the operational stress, and the chance that any wing will not be able to do so is extremely small.

Similar standards now being developed for the equipment components of aircraft and missiles, can provide safety margins with respect to functional time which will accommodate even the production variables associated with complex equipment. At the same time the art of controlling variations is also being improved. Thus, it is gradually becoming possible to design and produce equipment items so strong and so nearly identical from part to part, that there is only a small chance that any item will fail to function per specification for the required period of time.

This advance is being achieved largely through the development of time records which, every day, provide new knowledge of equipment characteristics under various operating conditions.

A few years ago very little failure rate data was available on even the most commonly used components. Today, the data developed through time recording tells us with good confidence what failure rate to expect from a large number of components. Both the number of components and the range of conditions on which data is available are rapidly expanding as more and more data is collected, correlated and analyzed.

The task of collecting information will never be completed. New components, and design and production improvements on old components, obsolete the data almost as fast as it is acquired. Even without changes the task would be very large, because of the large number of different environmental conditions and different system applications which affect component failure rates.

Still it is very helpful to the designer of an electronic "black box", for instance, to be able to know today that the average failure rate for transistors, as a class, when operated at rated loads, is .03 per 1000 component hours; (3 failures in 100,000 component hours.)

Since many of us in the aircraft and missile industry are thinking several years into the future, it is also helpful to know that, based upon expected improvements in design and improved control over production variables, the failure rate predicted for transistors produced in 1965 is only .002 per 1000 component hours; (2 failures in 1,000,000 component hours.)

This type of data, available today for diodes, resistors, capacitors and many other components, as well as for transistors, is not just a benefit, it is an absolute necessity to the designer faced with a requirement for equipment which will have a .999 probability of survival in any given hour of operation.

Without this data, it is possible that the engineer will release designs which, although they could be reliable if absolute control of production variables were possible, will, because of our inability to exercise complete control, predestine the resulting equipment to an unsatisfactory level of reliability.

Inevitably, when the lack of reliability becomes apparent, Design will blame Quality Control, and Quality Control will blame Design.

This is like a football team where the linemen are not told what plays the backfield is running. After losing the game it is inevitable, people being what they are, that the line will blame the backfield, and the backs will blame the line.

Reliable equipment is produced when Design, Production and Quality Control act as a Team. Production and Quality Control must recognize the need in Design to save weight and increase performance, and make every possible effort to improve the control over production variations in workmanship and materials. Most important of all, the degree of control achieved must be quantitatively measured, and reported to Design.

Design then, will be able to recognize the limitations of controls, rather than just complain about them, and can provide safety margins to accommodate the uncontrollable differences between successive production parts.

Data on the distribution of observed failure times about the average time per failure, taken under actual or simulated environmental conditions and subjected to proper statistical treatment, is one good measure of the difference between successive production parts, and hence, a measure of the control of production variables.

The planning of a good time recording system is not a simple matter. Time records are relatively useless without simultaneous recording of failures. This implies that "failure" must be defined for each component on which records are kept. Equipment seldom runs continuously; usually it is on and off intermittently, starting many times during its life. Since wear-out due to starts is often more severe than wear due to running, some method must be found for counting the number of starts and providing the proper statistical treatment of the combined start/run data.

The physical process of acquiring the raw data is often troublesome. Both clocks or counters and human observers have their limitations and attendant problems. Finally, the recording of time data is not inexpensive. A great deal of planning is necessary to produce an effective program, and a considerable amount of manpower may be absorbed in the acquisition and subsequent handling of the data.

Nevertheless, the benefits to be gained from a well planned program of time recording can far outweigh the trouble and expense. Time records can pin-point trouble areas quickly, and at the same time provide an indication of the most effective remedial action.

Time records permit a quantitative assessment of equipment reliability, which usually is not possible without time data, and also are very useful in tracing equipment reliability improvement. Intelligent analysis and application of time record data to engineering and quality control problems in the weapon system industry is today making it possible to produce more reliable initial designs, better control of quality, and earlier development of more effective weapons than ever before, with producer/consumer benefits that perhaps can be counted only in terms of national survival.

PRACTICAL APPLICATIONS OF EVOLUTIONARY OPERATION

John C. Whitwell
Department of Chemical Engineering
Princeton University

Unless there is a real need for process improvement, there are no practical applications of evolutionary operation. Therefore the first objective of this paper is to confirm this need. Subsequently solutions to some problems which have been encountered in practice will be discussed.

Need

Two major classifications exist, new and old plants. In the old it is theoretically possible for the plant personnel to have operated to an optimum empirically. However, there are two reasons for such optimization to be rare, despite the fact that a multiple, single-factor experimental approach will eventually find an optimum on most surfaces. First, plant runs to determine the effect of a given variable are conducted in the presence of large "noise", experimental error due to uncontrolled, or uncontrollable variables. Thus runs conducted at values of a variable which would properly be definitive if the response were obtained without error will be frequently non-definitive, if not thoroughly misleading, when conducted in the presence of large error. Second, plant runs are usually conducted to optimize some response measured at the unit. As will be noted later, this type of response may not optimize in the true sense.

As for new plants, it might well be argued that the advances in scientific knowledge allow much better development and design than in the past and that optimization can be effected before erection. A book entitled "Scale-up in Practice" (2) will reveal the weaknesses in such an argument. Despite the risks of quoting out of context, the following excerpts from this source are noted: "As the physical processes become more complex, ..., scale-up is still possible in theory but may require fairly extensive data on the smaller scale equipment. When chemical reactions are present the situation becomes sufficiently complex that true similitude is no longer possible in the general case ..... Finally, complications due to imperfect flow conditions .... must be understood in detail .... in order to permit scale-up. ... Such knowledge is not available at the present time, however." And in relation to the

pilot unit, "this unit was creating more problems than it was solving... Experimental work is loaded with traps.... Problems which shouldn't exist are often created in the experimentation... The importance or even the possible existence of the plant problem was not anticipated.... (Among other things) impurities literally get lost in small scale experimentation."

While this is not the complete story, it is sufficient to indicate that all the problems are not solved before the process goes on stream. There is a problem of optimization in most plants which can only be solved at the unit.

Method

There is no intention of reviewing here the method of EVOP. This audience will certainly be familiar with it, and with its potential of becoming the most important contribution to control of industrial production since Statistical Quality Control.

Like SQC, EVOP has such beautiful simplicity that problems in its application apparently should be minor, if not completely non-existent. Like any sound analytical method, however, its final value is dependent on the interpretation.

The emphasis here will therefore be placed upon interpretation and problems, with knowledge of the method assumed. Currently, the best reference on method is still contained in the original article (1).

Rules of the Game

The primary rule is very simple: DON'T DISTURB PRODUCTION. Those who give advice are not responsible for production; their suggestions are therefore entirely subordinate to the desires and decisions of those who are. Whenever a disagreement exists, no matter how major or minor, the final decision rests with the operating personnel. If, for example, there is a suspicion that the range of variables is being made too restrictive, the restricted range must still be accepted. The process of EVOP will prove who was right. If the operating personnel were correct in advocating small changes, the advisor will be glad that he did not assume responsibility -- even if he could have -- for changes of a destructive size. If larger changes could have been made, the early stages of the EVOP will show the low influence of the variable limits chosen, and these limits may be expanded.

Therefore to the rule, add the corollary: INTERPRETATION OF EFFECTS WHICH WILL DISTURB PRODUCTION RESTS SOLELY WITH PRODUCTION PERSONNEL. And finally, include the axiom: IF PRODUCTION IS ADVERSELY AFFECTED, YOU WILL NEVER HAVE ANOTHER CHANCE AT THIS PROCESS.

The game must be played solely under production's rules. The sequence, not the character, of the plays is the area in which personnel outside of production may be helpful.

Number of Variables

There is nothing inherent in first order designs which will constitute a theoretical limit on the number of variables which may be treated simultaneously. However, a natural restriction is imposed when one presumes that the interpretation of results should be performed in the plant. There is no intention to imply that plant technical staffs are less bright than other chemical engineers or statisticians. In fact, when one sees the problems faced daily by these people, one is likely to concede that they must be brighter than most to succeed. However, everyone in the unit can profit from a chance to see the progress being made, and not all of these people have the benefit of a technical training; thus to expect them to think in hyperdimensional space would be both presumptious and unreasonable. Moreover, speed of interpretation, and calculations involving uncomplicated techniques, are desirable at plant locations. Within this limitation, therefore, there is no recourse but to limit design points to three-dimensional space -- i.e., usually cubes.

With cubes, there are four systems which can be investigated, one, two, three and four variable designs. Exclude the one-variable design as a very special case required infrequently. Two variables can be most easily handled and interpreted visually in the square with a center point and are the logical choice in early applications. Three variables offer two possibilities, the full cube with eight points and one or two center points and the half-replicate with one center point. Admittedly this last does not estimate interactions, safety valves in the case that local region of the surface investigated is not truly linear. But the linear response will usually be true due to the restriction of the ranges of variables to meet the primary rule of production. And the safety valve is still provided by the center point. The half-replicate is then an excellent design for three

variables. The full cube, with half of the design for one level of the fourth variable and the other half for the other level is also reasonable, although the interpretation of the results in three dimensional space is now becoming complicated.

Undoubtedly many, if not all, processes involve more than four variables which should be investigated. Thus limitation to a practical maximum of three, and an ultimate limit of four at a time, suggests that some of the variables will never be included in the same phase. While this is basically a true statement, it ignores the possibility of permuting the arrangements of variables. True, the size of the total program, and the time to cover all is getting large. But the process is being run for production anyway; as long as it continues it may as well provide information, so a long program is reasonable. The argument against this philosophy is the time to reach the ultimate optimum. It might be argued that it is reduced by considering all the potential variables coincidentally in the initial program.

Consider this hypothesis. The extensive design covering all variables will have to be repeated a number of times to build up precision, as does the simpler EVOP design, and this replication will require more time than a single phase of EVOP since the number of runs in the basic design will be greater. During this time no information will be obtained regarding moves which might effect improvement, while in the same interval the initial large gains will already have been realized in the EVOP program. Each subsequent phase will have to be conducted in the same manner, carrying the relatively unimportant variables, with their inherent delaying action. There seems then to be no great fallacy in the sequential nature of consecutive small programs to cover the entire list of variables.

The Response

In a procedure for true optimization, there is no choice of main response to be made. It is obvious, and invariable-dollars. Unfortunately dollars are not usually measured directly at the unit while something like percent conversion or rate of production is. The temptation to EVOP on some such "natural response" is great and must not be succumbed to without carefully proving it equivalent to dollars. Rarely can it be expected that the natural response will tell the whole story.

As an example of the wrong response, consider the use of percent conversion, with all factors regarding productivity being ignored. Since the rate of production can be changed either positively or negatively while percent conversion is increasing, optimization of percent conversion leaves one completely ignorant of the total income, or profit, picture.

Similarly, if productivity only is measured, the relative consumption of raw materials may augment or counteract the gain, and a completely erroneous picture of the economic problem may again be obtained.

However, routine evaluation of a cost function may not be a guarantee of success either. Take the case where changes in raw material requirements and other operating charges were properly interpreted into cost per pound of product. The accounting department reports that, because of internal and external sales complications, it is impossible to state a single sales price. So, the EVOP reports improvement in terms of cost per pound at the standard production rate. So far so good, since a decided improvement is indicated. However the gain is relatively small, $75,000 per year on a large operation. And the figure does not account for a sizable change in production rate, since it is argued that the new production could be sold at cost. This argument is patently ridiculous. If product is transferred internally at cost, and if the cost has been reduced, some part of the organization is going to profit. If it is sold externally at the standard figure, the profit will show directly. Several sales prices in the normal range were assumed. The dollars response surface did not change in character with these different figures; it only moved up or down with the figure chosen. Improvement effected by EVOP, based on the sales price thought to be the best guess, then became $300,000 per year, for the previously estimated $75,000 case, in an actual application.

Qualifying Responses

Although dollars are undoubtedly the only true response, one must be very careful to avoid assessing improvement in dollars for an inferior product. Thus it is customary to record several responses, giving first importance to dollars, but following it with quality responses. Here, for example, include such things as color, pigment strength, amount of impurity, fluidity, etc., as guarantees that the product is at least as good as previously produced, and as a means of assessing the improvement in dollars value for

the main response. The number of these qualifying responses may vary from none as in the case of a highly purified fine chemical, to a dozen as with a product sold under a variety of specifications, and not susceptible to separate refining; for example, plastics.

Use of Existing Data

The question of the use of existing data frequently arises. Particularly there is the question as to whether existing data may not provide an estimate of error for use in assessing significant effects at the end of the first cycle, where the EVOP procedure has not yet produced its own estimate of the error. To my knowledge, no advocate of EVOP has ever suggested that such a procedure is fallacious. In fact, the original paper suggests the possibility.

There are, however, reservations. For example, one usually does not expect to experience strong effects in the first cycle unless the process is so new, or has been so badly run in the past that the gains are huge, so huge that benefits of replication as a means of increasing precision are not needed.

Second, it is unlikely that the variance has been obtained directly for the main response, dollars. The cost function may be written in terms of the other variables for which variances have been estimated. But in attempting to synthesize the dollars variance from these estimates, the experimenter should be aware of the pitfalls potential in the calculation of variance of a non-additive function.

As for use of existing data to substitute for EVOP, the answer seems to be simply NO. Even if all the variables enter adequately for determining their contributions in a multiple regression approach - a situation which will rarely exist, since the most suspect variables will probably be closely controlled -- the unbalanced design represented by the plant data will produce biases in the coefficients and the equation is essentially useless outside of the range of the variables encountered. Since the object would be to move the process to a new position, extrapolation can lead only to disaster. This problem of undesigned data seems to be inherent in at least one automatic optimizing procedure. The pitfalls of multiple regression must then be built into the instrument. Since optimizing involves first, an evaluation of the surface and second, extrapolation beyond the region of the data, these pitfalls can become serious.

Confidence Limits; Interpretation and Use

There is an unfortunate tendency to look upon confidence limits as absolute rules while they are, in fact, guides. The confidence limit is only intended to aid the user in assessing the magnitude of the effect measured in comparison to the uncertainty inherent in his particular program. There is nothing magic about the 95 percent level; it happens to be convenient because of the way in which statistical tables are reported and because odds of 20-to-1 against an incorrect statement are reasonably safe in most problems. But the failure to attain "significance" at the 5 percent level (the 95 percent confidence limits) certainly is not proof that no effect exists. If the effect were "significant" at the 80 percent level, the odds against no effect are still 4-to-1. At this level, a few more runs, and consequently higher precision, might well raise the same size effect to, or above, the 95% limits.

In addition one cannot overlook the fact that the percentage level calculated is based upon assumption of the normal distribution and the conduct of the program so that the assumption will obtain. It is true that the Central Limit Theorem does tend to assure normality as the limit when enough forms of random variation are mixed. But there is no guarantee that these conditions pertain perfectly in any program, and departure will affect the absolute values of the limits.

All these remarks lead to the conclusion that confidence limits are a reasonable guide, but only a guide, and too strict an adherence to the "rules" may well result in real trouble.

As an example of the misuse of confidence limits, take the two-variable EVOP where one variable was clearly producing an effect, while another was not "significant". A change in the "significant variable" was made in the direction indicated, and one of the levels of the "insignificant variable" was retained. In the case in point, the low level of the latter was retained and the results from the next phase were poor. Investigation of the previous phase disclosed the fact that while the second variable did not attain a significant level, it had a strong negative effect. The best gamble would have been its higher level, not the lower level chosen. Further investigation indicated that the reason that this variable had not shown a "significant effect" was largely due to a conservative decision with

regard to the range over which it was varied. Thus, while X_1 was clearly important, X_2 had been so conservatively scaled that the same effect on response, pound for pound, as found with X_1, had been marked "non-significant". It should have been treated in the analysis as equally important.

It should be thoroughly clear to the unit heads that the method of calculating confidence limits requires an understanding of the simple principles of linear combinations. Then, when a new design is undertaken, the estimated σ will not be improperly calculated. This principle of linear combinations is not complicated, and can readily be explained to technical personnel whose education in statistical methods has been neither extensive nor even elementary.

Moves to New Phases

At the end of every cycle a decision should be made with regard to the next series of runs. A variety of decisions may be made. Some of them might be:

a. continue through another cycle since no effects are strong enough to warrant a firm decision.

b. expand the range of one or more variables to aid in sharpening the contrasts.

c. shift the entire pattern to a new location.

d. do exploratory runs along the suggested path.

Most of the actions are comparatively clear. However, there is always the problem of the extent of the change, as in c and d. Exploratory runs are most helpful in establishing location of the next phase. There is no point in arbitrarily moving the whole design one unit when it could have been moved three or four units without disturbing production adversely. Thus small changes in a desired direction, replicated against the center point of the previous design seem to be a logical procedure and particularly whenever it is suspected that more than one unit change will be profitable.

Do not overlook the possibility that an "insignificant" variable may be important (see previous remarks on confidence limits), having failed to show a real effect only because the phase has not been run long enough or because the original

scaling handicapped the design by providing very small increments of this variable. Considerations of this type will allow one to draw judgments of the type:

a. include the effect of this variable in the trial direction.

b. run the current phase a little longer to show this effect more strongly, if it truly does exist.

c. use wider range on suspected variable.

d. in the design of the new phase, with new variables - if this is to be the case - take the more favorable position of the non-significant variable; if truly non-significant, this move will not hurt; if truly significant, this choice will eventually prove to have been wise.

No remarks about changes from phase to phase would be complete without the precautionary remark that the basic rule of EVOP -- don't disturb production -- must still be a primary consideration.

Interactions The usual practice is to omit discussion of the interaction term. This is fundamentally sound, since the design is so restricted by the normally small changes in variables that response curvature may not be detected. However, in rare instances a strong interaction may appear. In these cases the form of the curvilinearity, and the reason for its existence should be investigated. The form may readily be detected in the simple designs as ridges, rising ridges, hills, or even saddles. (The last would normally be suspect, suggesting that one of the other forms might well develop with more information and sharper contrasts.) But, regardless of the form predicted, great assistance is afforded in the selection of a direction.

This part of the problem is not the one of most interest. The one of prime importance is the possibility that, particularly in the early stages of a phase, one bad run may lead one into conclusions regarding curvature which cannot be substantiated later. Such cases exist. Examination of individual responses, and analysis of the residuals has, in specific instances, prevented some very erroneous interpretations.

Responsibility for the Calculations

The question frequently arises as to the proper place for the performance of the simple, but necessary, calculations. There really seems to be little choice. The operation of the unit is the responsibility of production personnel. Therefore calculations leading to decisions in the plant belong to the plant also. A full understanding of all matters pertaining to the decisions is essential for intelligent conclusions.

Moreover, if the calculations are performed by a service department, interference with the regular duties of this department must result. The duties of a statistical department include assistance in establishing new EVOP programs and performance of other services of a design nature. Both the design function, and the number of EVOP's which can be supported will be hampered if responsibility for calculations also rest with this group. Moreover, the "news" from the latest plant runs reaches the process "cold", when it should be available for immediate consideration.

In cases where the statistical department does the calculations, it would appear that the situation has arisen not so much from a failure to appreciate the preceding points, but rather from a desire to sell the EVOP program. The units are asked "Do you wish to do the calculations, or would you rather have us do them and report results for your consideration?" The answer to this question is obvious. No one is looking for unnecessary work. There is no need to go to this extent to sell EVOP. The unit head should be told that the statistical department will aid in setting up the EVOP, will provide form sheets for the calculations, and will check procedures in the first few cycles. After that, the unit is on its own except for the occasional meetings of an EVOP committee, on which statistics will unquestionably be represented.

Restrictions imposed by Process

It is not unusual in an old process to find that everyone believes that it could be operated at a better set of conditions, but a set which is unattainable due to equipment limitations. There is no question as to whether a change would be profitable. As a case in point, take a process in which the temperature is limited by the heating system. A higher temperature could be obtained only by installing a Dowtherm unit at a cost of $300,000. Is there any excuse

for running an EVOP to predict what is surely a known fact?

In the example chosen, there was no way to estimate, with any reasonable precision, the savings which might accrue with the Dowtherm unit. The EVOP was run, and it could easily be shown that the predicted saving per unit of temperature change would be $300,000. The number of the arbitrary units of temperature which might be expected to be attained is, of course, not estimated by the first order design. However, experienced technical personnel on this unit could make very shrewd guesses and it seemed clear that several units of change might be a conservative guess. Thus the EVOP, by providing good cost figures for the predicted change, afforded a means of estimating the pay-out time for the new unit.

This story should be concluded with the statement that the change was made, and that the savings were as predicted. Unfortunately, this is not the case. This unit was so old that it was considered obsolete and not worth the change. The EVOP was not lost, however, for the information was invaluable in the design of new units and for information regarding operation of newer units at other plants of the same company.

An Impatient Management

A few successes in the initial cycles of the first few EVOP's may easily become embarrassing. In instituting EVOP it is wise to choose processes which one can believe capable of large improvement. If guessed correctly, and a large potential gain is available, a very few cycles will indicate desirable moves, which will easily pay off. Then one may be faced by expectations that all further programs will also produce results in 3 to 4 cycles. Obviously this is not the case. EVOP is designed to find small, heavily obscured, effects, and many cycles will be required in this situation before the changes are clear. It pays off quickly only when a unit has been operating most inefficiently, as already noted.

Impatience may be hard to combat. You may be told to forget the EVOP in this process and move on to richer fields. There are several alternatives. First, it may be possible to explain the true job of EVOP and thereby buy the needed time and effort. Second, it may be possible to persuade management and the unit head to increase the range of the variables; changes made so far have not hurt production and

somewhat less conservative changes in the variables may then be reasonable. Third, by treating EVOP as a continuous mode of operation, essentially without cost when the calculations are performed at the unit, it can be shown that the continuation does not hurt anything and it may eventually help. In general, one of these approaches should buy the needed time when it is clear that no great cost or effort is involved.

All units do not necessarily agree that the cost of operating an EVOP program is negligible. In one plant where many successful EVOP programs have been performed, one department states that the evaluations required are more extensive than in standard practice, and therefore more costly. In some instances, the number of different responses to be measured is also increased. Under these circumstances they argue that the only excuse for EVOP is in the quick payoff. It does not seem to be easy to justify such conclusions. If the responses are important in the EVOP program, it would appear that they are required in normal operation. If eventual gains are available any program which reduces the responses measured appears to be poor economy. A system which produces sharper contrasts must aid, both in total improvement and in the time to attain it.

The Uncontrolled Variable

A completely uncontrolled variable of non-random character, such as variable feed stocks or declining catalyst efficiency, may well pose a gigantic problem. There seems little doubt that the proper solution is some form of continuous optimization, with variable input in a design which will make the effects as clear and unbiased as in the intermittent EVOP; responses are fed to a computer, with feedback to the variable control system. An EVOP application of this type is currently being studied jointly by the Statistical Techniques Research Group and the Chemical Engineering Department of Princeton University.

Other forms of automatic optimization equipment, some of which are familiar to the author, are rapidly appearing. Comment has already been made on one system. Others seem to have been treated on the relatively trivial basis of one variable, where many of the important problems posed by coincident changes in more than one variable are by-passed. Still others use a random choice of direction of move (3), reported to produce results as good as steepest ascent; since these depend upon comparisons of the result of one

run to that of the next as a criterion of improvement, and since they cannot obtain precision in this comparison due to exclusion of replication, there are obvious difficulties in obtaining a true optimum. Still others (3) use very expensive computers with systematic perturbations of the variables.

Time and experience will be required to evaluate these, as well as the automatic EVOP, both technically and economically. Certainly the cost of these systems will be appreciable, the ultimate costs varying with the particular application. Very careful study of potential gains will be required before committing any unit to the expense involved. In many instances the magnitude of the process being optimized may well be insufficient to justify the cost of the more ambitious methods. In fact, it seems certain that the most ambitious will have to be limited to very large installations.

During the period of development of the various automatic optimizing systems, some uninstrumented procedures for special cases of uncontrolled variables may be of interest. One instance is known in which feed stock varies outside of the control of the unit, but falls into distinct classifications within which variations may be considered to be random. Thus each feed stock is capable of being EVOP-ed separately. Parallel programs were established, each being run as its own raw material type became available. In such cases precautions are needed for the presence of block effects.

In another plant cases of declining catalysts are commonplace occurrences. Several approaches have been suggested. One is the use of inter-cycle effects as a measure of the decline of the catalyst, with the directions of change at each average level of catalyst activity being determined, and used for local improvement as well as for prediction of the starting points of EVOP's with the next charge of catalyst. Another is the institution of simple second order designs (such as the hexagon in place of the square for a two-variable design). Both the direction and the extent of the desired move would then be determined simultaneously. For success here the range of the variables must be sufficient for the second order coefficients to have strong effects; the operation should be near the optimum so that extrapolation of the local surface investigated need not be extensive. Yet a third is operation in perfectly standard fashion. Where large gains are to be obtained,

this has been entirely successful. In one case, advantage was taken of the fact that the catalyst efficiency was higher after the EVOP than it had been at the beginning. In normal operation the efficiency declines. Therefore in calculating the gain it was possible to include: gains in cost per pound; gains in productivity; gains attributable to increase in efficiency over a base of constant efficiency, or the efficiency at normal rate of decline.

Thus, although the conditions may not exactly fit the desired pattern for a conventional EVOP there is no reason to believe that a little ingenuity may not still produce a good part of the potential gains.

Summary and Conclusions

This paper was primarily designed to discuss EVOP in relation to specific applications. No attempt has been made to detail methods or calculations. Every point results directly from special features of EVOP programs, or interpretational and management problems which have been encountered in practice. They have been included here in the hope that others who advise or conduct such programs may find these experiences helpful in solving similar problems.

LITERATURE CITED

1. Box, G. E. P., J. Applied Statistics, 6,81 (1957)

2. Fleming, R., Editor, "Scale-up in Practice", Reinhold, New York, 1958.

3. Symposium, "Computer Control of Processing Units" 39th National Meeting, A.I.Ch.E., Atlantic City, 15-18 March 1959.

QUALITY ASSURANCE ENVIRONMENTAL TESTING

Commander Glenn Estes, Jr., U.S. Navy

There have been many thousands of words written and spoken on the subject of the reliability of complex weapons, the manner in which reliability can be achieved, the problems confronting the designers and manufacturers in attaining a high degree of reliability and the chaos which results when reliability objectives are not achieved. Some of the ideas I have heard expressed have discussed a manner in which to organize to achieve reliability. Other ideas and papers have been devoted to the presentation of an idea on establishing monitoring points for the determination of reliability. These expressions represented varying degrees of practicality, but all of them had one objective in mind and that was to achieve, and to help others to achieve, reliability in complex weapons systems in order to give an operational commander a high level of confidence in the use of the weapons placed at his disposal to aid him in achieving his tactical and operational objectives.

The Bureau of Ordnance has always practiced and believed in active government participation in the inspection and quality control practices used by its contractors, particularly in the production of complex weapons such as guided missiles. This belief is brought about by a number of considerations. The first consideration is the need for placing into the Fleets and into stockpiles reliable weapons which can be called upon and used at a moment's notice. Another consideration is that of cost. Prices are rising and have been rising for some time in the past. The achievement of high reliability would enable the Bureau of Ordnance to place smaller numbers of complex weapons into the Fleets and stockpile, thereby realizing considerable savings in both dollars and time. There are seldom, if ever, enough of these two commodities available to us.

Still another consideration which the Bureau of Ordnance must take into account is the complexity of design and manufacture attendant to the production of modern weapons. This complexity restricts the amount of maintenance which can be provided to these weapons in the forward areas and on board ship. The limited amount of maintenance which we can accomplish is usually restricted to the replacement of certain removable spare parts or of making a limited number of adjustments to the missile or test equipment. Therefore, we must design and build into these complex weapons a high degree of reliability in order to alleviate this problem. While it is most certainly true that reliability cannot be inspected into a product, it is equally true that proper and thorough inspection techniques can confirm the presence or absence of the reliability initially designed into the product.

The safety of personnel and equipment is yet another point which must be taken into consideration by the Bureau of Ordnance. The malfunction of a weapon within the restricted confines of a ship, for example, not only jeopardizes the weapon and its immediate operators, but also jeopardizes the ship and all of its personnel as well.

One of the most important considerations the Bureau of Ordnance must recognize in guided missile operations is the finality of the use of that guided missile--the "one-shot" characteristic. Unlike an airplane, which is most cases lands back aboard ship after completing a mission, a missile is launched and either successfully completes its mission or it doesn't. There is no further opportunity to talk to its "pilot" and determine what went wrong in case of malfunction. All of these "conversations" must be held prior to launch through the medium of test equipment. At sea, limited numbers of missiles are available to a ship or operational commander for use in air defense. This fact, coupled with the "one-shot" characteristic, demand that our missiles have the highest economically achievable degree of reliability.

The Bureau of Ordnance has always believed and practiced the belief that national security is an "ALL HANDS" job. The Chief of the Bureau of Ordnance has a responsibility to the Fleet to provide ordnance equipment of the highest possible quality and reliability. At the same time, the specifications under which this ordnance is produced cannot call out such reliability requirements which force it to be unreasonably costly. I believe that it is the function of any quality control organization to strike a sensible balance between quality and cost. In striking this balance, the quality control organization must have complete cooperation from all other organizational elements in the production cycle of the equipment.

If then, national security is an "ALL HANDS" job, it follows that there must be real cooperation between industry and government in getting this job done. One of the ways in which this cooperation is practiced between the Bureau of Ordnance and its contractors is in the evaluation of contractors' inspection systems. A part of every Bureau of Ordnance contract states that the contractor must maintain an inspection system acceptable to the government. The Bureau of Ordnance determines the degree of acceptability by performing evaluations of the contractor's inspection system in accordance with the provisions of NAVORD INSTRUCTION 4355.22. The contractor participates fully in the evaluation and, in most cases, undertakes to correct the deficiencies which are pointed out by it.

Another way in which cooperation is manifested is in the evaluation of production acceptance test equipment, which, due to its complexity and cost, is used concurrently by both the contractor and government inspectors. All test procedures, test plans, etc., covering this type of evaluation are prepared by the contractor and submitted to the Navy for approval. The test equipment is then evaluated by the contractor and the Navy working together to determine its effectiveness in accepting or rejecting good or bad production material.

Still another way in which cooperation is practiced between the Bureau of Ordnance and its contractors is in the application of a Coordinated Industry and Government Acceptance Inspection Plan (CIGAIP). This is a plan developed by the Navy's Bureau of Ordnance whereby government inspection is reduced in scope as the contractor demonstrates his ability and willingness to perform and document his inspection procedures and results in accordance with government requirements. Where this plan has been implemented it has been most

successful and mutually beneficial. The details of this operation are spelled out in BUORD QC DIVISION INSTRUCTION 4355.10.

Now, I will describe one of the specific steps which has been taken in Navy missile programs in order to achieve greater reliability in our guided missiles. This program is entitled "QUALITY ASSURANCE ENVIRONMENTAL TESTING" (QAET), and is presently in use in the TERRIER missile production program. It is planned for use in the TARTAR and TALOS production programs, also. Essentially, QAET is designed to accomplish three things:

1. a continuing measure of product quality under environments of temperature, vibration and humidity

2. the basis for corrective action, and,

3. information to the customer on precautions to be observed in the use of the product.

A six month environmental test program is performed on all outside purchased parts and components which go into the TERRIER missile. In the parts testing program, samples are selected at least once from incoming lots from each source and subjected to the environments under which the individual part showed sensitivity during prior testing, i.e., in the parts qualification, pre-production or design evaluation testing programs. The sample of each individual part purchased from outside contractors is subjected to at least two environments. Sample sizes are determined according to MIL STD 105A, Appendix on Expensive Testing, to give an AQL equal to 1.0.

In the components QAET program, each monthly shipment lot is sampled and each sample is subjected to an average of three environmental tests. The environments selected are again those environments under which the component showed a border-line capability during pre-production, qualification and design evaluation testing or those environments which are considered to be the most critical relative to the system application of the component. As stated before, testing continues for six months, giving a total of six sets of tests to each type of component bought by the prime contractor from each outside source. Sample sizes are again dictated by the Expensive Testing Appendix of MIL STD 105A to give an AQL equal to 2.5.

The final tests performed under the QAET program are made on eight Guidance and Control groups randomly selected from monthly production starting with the first months production and continuing for six months. These tests are non-destructive in nature and the test environments are again those under which the Guidance and Control groups showed the greatest susceptibility during prior test programs.

Test data and reports are furnished by the contractor to the local Naval Inspector of Ordnance and to the Bureau of Ordnance. It must be pointed out that this kind of a test program calls for the finest kind of cooperation between the contractor and the Navy. When failures occur, the contractor's decisions and actions must be dictated by the application of the best available engineering judgment based on consideration of the governing specification and the

environmental conditions to which the missile is expected to be subjected during its Fleet operational use. The Bureau of Ordnance or the local Naval Inspector of Ordnance is always consulted in the determination of these environmental conditions.

From the test results generated thus far in the TERRIER program both the contractor and Navy are very satisfied with the performance of the program and the results which have been achieved. Many defective parts have been isolated and replaced. In some cases, individual sub-contractors have been required to rework one hundred percent of the material which he furnished. Up to this point, every sub-contractor concerned has been most anxious to replace defective material after he has been satisfied that his material had been subjected to fair and impartial testing. This has been most encouraging, in my opinion.

Now what do we do with all this information after we have once accumulated it? It has been my feeling for some time that there is not enough interchange of information between missile programs. I know this is true in the Navy programs, and I suspect that it is equally true in the other services. In pursuit of this objective of information interchange, the Bureau of Ordnance has placed a task at the Naval Ordnance Laboratory, Corona, California. This task assignment calls for the collection, reduction and analysis of the data generated in the TERRIER QAET program; the assembly and presentation of the results in some sort of highly readable, handbook form. These handbooks will then be distributed to designers and engineers working in other Navy missile programs. In this manner we hope to be able to save time, money, manpower and material in presenting reliability information on pieces, parts and components common to Navy missiles. Our ultimate objective is to collect, reduce, analyze and present similar data generated in all test and surveillance programs. I fully realize the immensity and complexity of this task, but I also recognize the need for the successful completion of such a task.

The thoughts, ideas and opinions I have expressed are neither new nor startling. They can best be summed up by saying that I believe that great strides can be made by the industry/government team toward the achievement of reliability goals by free interchange of reliability information; by application of the principles of hard work, diligence and cooperation; by mutual understanding and appreciation of the problems which are peculiar to government and industry; and by continuing to perform a maximum effort to research in reliability problems.

FORD'S QUALITY AUDIT PROGRAM

John A. Fournier, Manager, Chassis Quality
Ford Motor Company

Since the end of World War II we have witnessed within the automotive industry an ever growing complexity of car and truck designs, a wider variety of factory installed options, frequent styling changes and shorter design, tooling, and manufacturing lead time. In addition, as the automotive market turned to a buyer's market after World War II, customers grew to expect better and better quality.

One of the most challenging problems facing the automotive industry today is the control of product quality. Customer acceptance is vital to the success and growth of any business and product quality is a fundamental requirement. Probably one of the most difficult problems for our entire industry has been the maintaining of a consistently high standard of quality in all units at all plants at all times. Some may believe this is a characteristic of high volume assembly operation. If it is, I am sure our customers will not accept or like the results.

In any highly competitive business such as ours the manufacturers must constantly seek ways to add more value to their products - more value than their competitors are able to provide. Improved style and better engineering are two fundamental ways to accomplish this. Others include being more efficient in manufacturing, cost control and product quality. The customer, of course, is the beneficiary of all these efforts.

Efforts to achieve better quality have taken several shapes and forms. Generally we have relied most heavily on in-plant inspection. Our studies showed, however, that in-process inspections had very definite limitations. To produce between six and seven thousand vehicles daily at our assembly plants involves the assembly of approximately thirty million parts. It is impossible to inspect all of these parts in assembly operations 100%. Of course we all know that you cannot inspect quality into a product but it must be built into the vehicle from the start. It was our conclusion that if we were to achieve the quality edge that we sought we would have to undertake a fresh approach.

In looking about us it was evident that we could tell how many cars we built daily because it's easy to count cars. We could also measure our cost performance because it was easy to count dollars. However, it was very clear there was no objective method of evaluating quality performance. This lack of a quality measure prevented our plant and division management from placing cost, production and quality in their proper relationship on a daily basis. Without this relationship proper stimulation to quality aspects could never be expected.

Some way had to be found to provide a uniform objective measure of quality which would accurately reflect customer requirements. What we required was a standard method of measuring quality that could be applied across the country and to all types of units, including trucks. It had to be accurate, uniform, and simple. Further, the results had to be measurable on a scoring system which would provide plant and divisional management the information they required to evaluate performance.

Within our assembly operations the system must identify our weaknesses and trace them back directly to the point in the process where they started. In total what we required was a system which would allow us to measure quality like we count dollars or cars and allow us to manage quality just as we manage production and finance.

After some experimentation and study we developed a system which we believe meets these requirements. We have chosen to call it the Ford Quality Audit Program. I would like to tell you in more detail how the system works, what we believe it has accomplished and where we stand today.

At each of our plants there is an audit team whose assignment is to audit cars that have been approved by the plant inspection system. The audit team is comprised of salaried employees who report directly to the plant quality control manager. Their job is to make a random selection of cars ready for delivery to dealers and put each car through a comprehensive examination. Each complete audit requires a visual or operational check of some 1500 different items. A complete car audit requires approximately three hours of intensive scrutiny. The problems indicated on each car audited are classified and noted on special forms.

The Body Visual Audit Record is used to detail all body appearance items such as those related to metal finish, solder joints, paint, door, hood and deck lid fits, interior trim and paint, security of weatherstrip and wind cords and glass and window operation.

The Body Water Test Record pinpoints all leaks detected from the exposure of cars in specially designed water test booths.

The Chassis Visual Audit Record is used to record results from the following checks: Instruments and controls such as electrical switches and indicators, operation of accessories, brake, clutch and transmission controls, under hood inspection of such items as belt alignment, carburetor, spark plug connections, transmission linkage, air cleaner, coolant level, brake connections, wire routings, etc. The reverse side of this record is used for the under frame audit of frame welds, steering linkage, suspension, fuel and brake lines, shock absorbers, axle and transmission assemblies, fuel tank and many others.

The Driving and Measurement Audit is used to record performance characteristics such as engine starting, idle, acceleration, handling qualities, noise level, operations of transmissions, brakes, clutch, etc. The driving portion of the audit involves a road test over a prescribed road under varying driving conditions.

The measurement portion of this audit, which is recorded on the reverse side, involves a torque check on all critical points of attachment. It is not just a question of seeing that the bolts are tight but whether or not they are within the specified limitations - neither too tight nor too loose. Other measurements are taken of the caster and camber settings on the front wheels, the amount of play in the brake pedal, accelerator pedal height, headlight aim and many others.

The audit system is executed according to uniform procedures and standards which leaves little room for variance. In this connection it

is well to mention that the plant auditors are checked by other auditors periodically. On a regular scheduled basis auditors from Dearborn show up at each assembly plant to audit cars and trucks. If the rating given by the General Office auditor and the plant auditors are the same we know the program is on target. Obviously in such a program constant vigilance and training is necessary.

At this point it is appropriate to mention that each defect has a rating, a uniform rating which is used throughout the Ford Division. For example, a condition which impairs the quality of the vehicle is rated at 2 points; on the other hand a slight imperfection, that is an out-of-standard condition - is incidental - counts only .2 of a point. A major condition counts 10 points. The total possible defective conditions of a vehicle have been classified in five different evaluations similar to the three examples given. This means we have established a numerical value for each defect and the numerical value is the same for all plants within the Ford Division.

Quality, like styling, is largely a matter of personal opinion. Unfortunately opinions can vary widely. It is always a temptation to use one's own opinion as being representative or typical of the average customer. Unfortunately, our customers have no accurate method of expressing their quality requirements - they too find it difficult to define quality. However, much can be learned from a negative customer reaction.

At the time the audit program was developed in 1957 it was necessary to establish our audit or quality standards based on our experience, judgment and warranty reports. It was apparent that we required a tool to constantly calibrate the quality standards applicable to Ford cars and trucks.

After giving consideration to such sources as warranty reports, market surveys, contacts with dealer personnel, it was concluded that all these sources lacked sufficient specific detail, timeliness and were too general in nature to meet our requirements.

A customer's request to repair a problem is a specific action resulting from his opinion of the product which can be readily identified and measured. Adoption of this principle led to a program which we have called "Dealer Repair Order Analysis" or D.R.O.A. which is the recording of the repairs made in Ford dealerships on a sample of production from each assembly plant.

One cannot question the validity of a problem which motivates the customer to relinquishing his car for repair. Obviously the customer's personal opinion is involved. However, this is an opinion which we must cater to regardless of our own personal convictions.

The measurement of production quality using valid standards will not create quality in itself. However, when daily audit ratings can be supported by field experience, management will invariably insist that they be met.

The development of a valid yardstick to measure product quality represents at least 50% of the task involved in controlling quality.

Most people resist being measured. Many will go to great lengths to avoid it or discredit it.

Now that we have seen how the cars are audited and the checks and balances that we have incorporated in the system for developing audit standards, let's see how the program works to improve quality. First, all the audit results from each plant are wired daily to the General Office in Dearborn. Since the inspections are uniform, the management is able to evaluate the quality rating of each of the plants. This only causes the various plant managers to work harder to get better quality.

However the system does provide more than just an overall quality report. The reports of the plant auditors as well as the Dearborn auditors are studied by the plant managers. Plant managers can pinpoint their problems and devise corrections. If, for example the auditor finds a cotter key missing this will show on his report because he will have to score it. The plant manager then can trace the defect right down the line to the person charged with installing the cotter key. Both engineering and supplier quality problems can be pinpointed for action.

In conclusion, we believe that the successful control of product quality involves three major requirements: (1) Quality standards closely calibrated to customer acceptance requirements; (2) continuous measurement of manufacturing performance against these standards; (3) the rapid identification of specific problems and the assignment of a responsibility for corrective action.

There is probably a fourth requirement which deals not with the standards and techniques employed but rather with the philosophy, desire and dedication of the management and others in the business having responsibility for the quality of the product. Ideally the management of quality can best be accomplished by the foreman and the operator. This is probably just as important an objective as being able to measure in understandable terms their quality performance. Most people have the desire and willingness to contribute their best toward achieving any reasonable goal. In the case of quality we must measure performance in terms which have meaning as far as our customers are concerned and be able to demonstrate correlation to our management and production people. It, therefore, is a fundamental responsibility of management to provide the right philosophy, interest, and leadership along with the proper tools. Under these conditions, the management of quality can be carried out by those best able to achieve results.

IMPLEMENTING S.Q.C. TECHNIQUES WITH MODERN DATA PROCESSING EQUIPMENT
- AN INTRODUCTION -

Douglas H. W. Allan
Metallurgical Statistical Analyst
American Iron and Steel Institute

Much has been written about the application of modern data processing equipment to business problems. There appears, however, to be one area of interest to the members of our Society that has been somewhat overlooked in this deluge of information. This area is the application of data processing equipment to the more basic problems encountered in Statistical Quality Control work.

In an attempt to explore this area, the Metals Technical Committee of A.S.Q.C. has sponsored a combined written and oral discussion represented by the following eight papers. Four equipment users from the metals industry have been chosen to comment on a different part of the following question:

a. You have three sets of data, each with a different card volume as shown below:

A - 100 cards
B - 1000 cards
C - 20,000 cards

b. How would you solve the following problems:

1. The first three moments of a frequency distribution and a plot of the distribution.
2. A simple correlation.
3. A five variable multiple correlation.
4. A thirty variable multiple correlation.

Four representatives of the manufacturers of the particular equipment employed by the users in solving their problems have been asked to comment on the solutions presented.

It is hoped that this discussion will highlight some of the problems encountered by both the equipment user and the equipment manufacturer in solving these basic Statistical Quality Control problems on modern data processing equipment.

Multiple Correlation for

PROCESSING QUALITY CONTROL DATA ON A DIGITAL COMPUTER

By Dr. M. J. Greaves

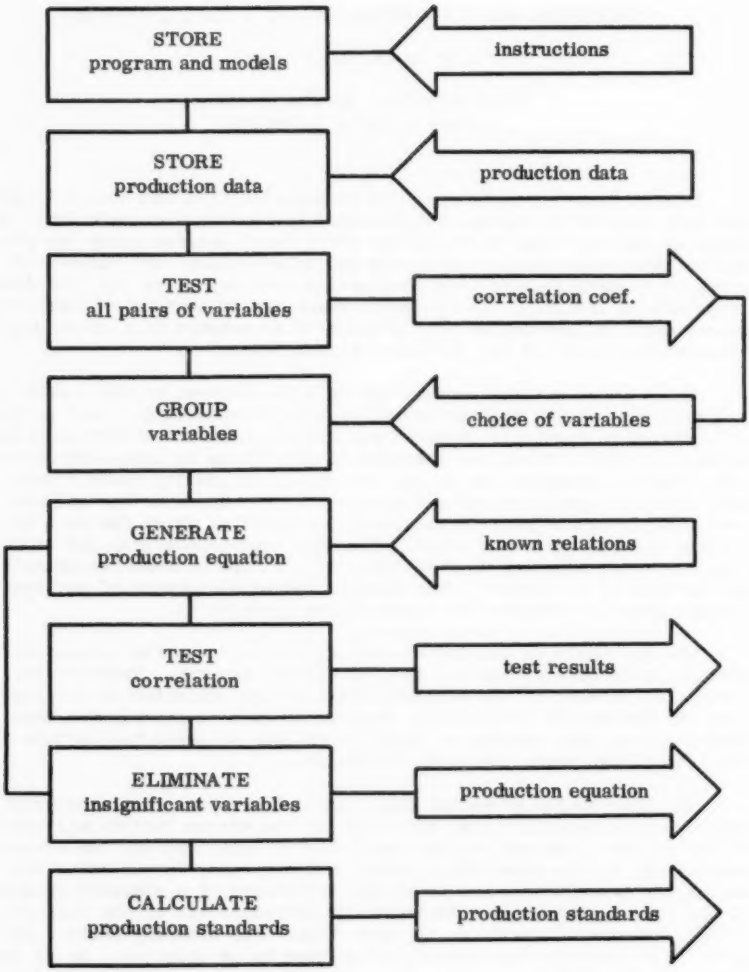
Chief Engineer, Metals Division
Arthur G. McKee & Company

Quality control techniques can be profitably implemented with modern data processing equipment programmed for multiple correlation. To discover the involved relationships which exist between cause and effect, necessitates collecting, organizing and interpreting vast amounts of data. The usefulness of this information will depend on its timeliness, scope and reliability. An automatic means for collection of compatible source data in combination with a properly programmed high speed digital computer meets all of the necessary requirements.

The highest quality product can only be achieved at the lowest cost by evaluating all of the factors which affect production. Much of this information is regularly gathered and filed but is not effectively used because of the difficulties inherent in converting it into useable form. Some computer programs now in use for analyzing quality control data make only straight-line correlations and do not discover the optimum levels. Such programs fail to ascertain the point at which the peak of production and quality is attained. The technique herein proposed differs from the usual approach in that the relationships between variables are not assumed to be linear. This permits the establishment of optimum levels and greatly enhances the value of the correlation.

The automatic collection of quality control data for mechanical processing is frequently the key to a successful program. However, the form in which most of the available data is now collected is not convenient for mechanical processing. Nothing seems more incongruous than hand-punching vast amounts of handwritten data on cards before they are fed into a high speed electronic computer.

In order to fully realize the potential benefits from a computer system, it is necessary that the scope of the system include all phases of the project from the collection of source data, through the stages of processing, to the point where these data are ready for ultimate use. During the introductory and developmental stages of a computer program it may not be necessary or desirable to introduce all of the features of a fully integrated system at the same time. For example, basic information is frequently collected in a log book by an observer. As an interim expedient, this information is transcribed and manually punched on cards to introduce this information into the memory section of the computer. This manual transcription obviously represents a serious limitation to the full realization of the potential benefits of the computer system and may destroy the economic advantages inherent in a fully automated operation. It is now possible and, indeed, equipment is now in operation wherein all manner of information is regularly and automatically monitored and logged directly on a storage medium compatible with computer equipment.



PROCEDURE DIAGRAM

The proper planning of a data processing system should include and anticipate a fully automatic network. When a system is built up over a period of time, it is essential that plans be made for the ultimate layout in order that parts may be added in an orderly fashion. This is easily achieved if all of the components are assembled in a building block fashion and are compatible. This requires a serious decision as to what communication and storage means will be employed. Such a decision cannot be intelligently made without considering all of the factors, including permanent storage of records and the ultimate as well as the initial scope of the system in terms of the aims to be achieved and the benefits to be derived.

The best computer equipment is useless without a program to direct the instruction repertoire which is built into the machine. The program for processing quality control data by the method of multiple correlation involves operations with numbers and transformations of symbolic information. Each step and all situations must be fully anticipated and the necessary procedure spelled out in the form of detailed and specific instructions. When this program has been introduced into the machine, the stage is set to determine the mathematical relationships between cause and effect.

A sufficient amount of data is required to lend substance to a multiple correlation. For example, if there are 30 independent variables and if only linear correlations are sought, there must be more than 30 unique sets of data available. As a practical matter, several hundred sets would be desirable. A sufficient range of observation is necessary for the results to be meaningful. In any event, since this is an empirical process, the results can apply, with confidence, only to the range tested.

The production data are read into the computer as sets. These sets include measures of the quality, rate and cost of one unit of production and simultaneous observations of each of the variables which are believed to influence production. Since two physical quantities may be observed which measure the same function, it is necessary to select the best one to be used in subsequent multiple correlation. All pairs of variables are tested to discover where this condition exists. In order to provide a basis to select which of the two will be used, a correlation is run for each one with respect to each measure of production. Since it may be desirable to include certain variables with a low correlation, the results of these calculations are read out and examined by a skilled operator to select the best set of independent variables. To date there appears to be no satisfactory substitute for experience and judgment to make this selection. The program, however, provides the basis for quick and intelligent decisions.

A mathematical model is needed to express each relationship in order for the computer to operate on it. This model may be a linear relation or simply an equation of the form, $P = au$. Generally, however, a linear relationship may not be assumed, and an equation with higher terms must be used to achieve an acceptable approximation. For purposes of illustration, a cubic equation

$$P_u = a_1 u + a_2 u^2 + a_3 u^3 \qquad (1)$$

is used. However in some cases, where a better fit is required, higher order terms may be included. On one occasion the program had to be modified to include a sixth degree term to obtain the desired degree of correlation.

In physical, chemical and metallurgical processes it is common for two variables to have an inseparable effect on production. For example, in the sintering process the moisture requirements depend on the amount of recirculating sinter fines. The optimum moisture level is different for various levels of recirculation. Likewise the optimum level of recirculation is different for each moisture level. Where this condition exists it is necessary to use a mathematical model which has similar properties. An expression which has been found useful for this purpose is:

$$P_{vw} = b_1 v + b_2 w + b_3 vw + b_4 v^2 + b_5 w^2 + b_6 v^2 w + b_7 v w^2 + b_8 v^3 + b_9 w^3 \qquad (2)$$

Similarly an occasion may arise where three variables have an inseparable influence on production. Again referring to the sintering process, the three quantities, acidic oxides, basic oxides and fuel, have an interrelated influence on production. None of these quantities can be considered singly. The following mathematical model has been used to correlate the rate of production with these three variables.

$$P_{xyz} = c_1 x + c_2 y + c_3 z + c_4 xy + c_5 yz + c_6 zx + c_7 x^2 + c_8 y^2 + c_9 z^2 + c_{10} x^2 y + c_{11} y^2 z + c_{12} z^2 x + c_{13} xy^2 + c_{14} yz^2 + c_{15} zx^2 + c_{16} x^3 + c_{17} y^3 + c_{18} z^3 \ldots\ldots\ldots \qquad (3)$$

A known relationship may exist between production and some of the operational variables. This relationship is introduced into the computer ahead of the production data in order to reduce the number of unknowns. The computer program is arranged to take into account the effect of these variables on production. This procedure simplifies the problem, reduces computer time and eliminates the inaccuracies which would creep in if it were necessary to learn all of the relationships simultaneously. There is some hazard, however, unless a careful distinction is made between proven relationships and those which are generally accepted but not proven. Such relationships may not apply or they may have been established by direction without adequate verification. If an inaccurate or incorrect relationship is used, the entire correlation may suffer or, even worse, may be misleading. A good rule to follow is, when in doubt -- leave it out.

Multiple correlation makes possible the reduction of all of the observed data to a set of mathematical equations. These equations are called the production equations and together with statistical measures of confidence represent the data. The variables in the equations are

separated into groups, each containing the smallest possible number of factors which are closely related in their influence on production. The quality, rate and cost of production are each expressed by an algebraic equation of the form:

$$P = P_o + P_{rst} + P_u + P_{vw} + P_{xyz} \qquad (4)$$

Where:

rst = typical variables which have a known influence on production

u = a typical variable which has a separate but unknown influence on production

vw = an example of two variables which have an inseparable and unknown influence on production

xyz = an example of three variables which have an inseparable and unknown influence on production

P = quality, rate, or cost of production

P_o = constant

P_{rst} = the known relationship between production and the variables r, s and t

P_u = an assumed expression for the effect of the variable u on production. See equation (1)

P_{vw} = an assumed expression for the effect of the inseparable variables v and w on production. See equation (2)

P_{xyz} = an assumed expression for the effect of the inseparable variables x, y and z on production. See equation (3)

Values are computed for the constant, P_o, and the coefficients, a_n, b_n, and c_n, which will produce the smallest sum of the squares of the deviations:

$$\sum_{j=1}^{j=n} d_j^2 = d_1^2 + d_2^2 + \ldots + d_j^2 + \ldots + d_n^2 \qquad (5)$$

where:

$d_j = (P_j - P)$ = the deviation

P = the value given by the equation, and

P_j = the observed value

A statistical test is made for each coefficient to ascertain the level of significance. Where little or no correlation is found, it is concluded that the variable associated with the coefficient is not significant in influencing this particular measure of production and such coefficients and associated variables are deleted from this equation. The remaining terms are reprocessed by the computer to determine a new set of coefficients in the same manner as before. The above steps are repeated until an equation is obtained containing only the terms which have tested significance and adequately represent the observed data.

At this stage in the program the production equations express mathematically each of the dependent variables, quality, rate and cost of production, in terms of the independent variables which are regularly observed and reported in the production records. The effect of each of these variables on production can be determined by substituting values for the independent variables in the production equations.

The standard deviation for each dependent variable is then determined to provide a reliable guide for properly interpreting the results. The production equations are then used to ascertain the optimum values for those independent variables which are subject to the control of the operators.

By substituting these optimum values, plus those of independent variables not under the control of the operator, the production equations make the prediction of operating results possible under practically any set of conditions within the scope of available data. Production standards may thus be established. Finally all of the results are read out and printed on a suitable form.

The production equations cannot replace human intelligence and cannot be considered as a substitute for good judgment. However, they provide an intelligent person with a sound basis for practicing his judgment. These equations apply only within the range of variables that are used in their development and can only be as accurate as the source data and the operating acumen on which their development is predicated. However, it is believed that insofar as the observed data adequately describe the operation, these equations constitute a good summary in a useful form. The procedure diagram illustrates the foregoing as it would apply to a Burroughs 220. However, it is sufficiently general that it may be applied to any computer system.

By adapting the production equations to fit specific plant operations, the following information can be obtained from a multiple-correlation analysis: (1) The optimum value for each significant operating variable and (2) The effect of each of these variables on the quality, rate and cost of production. The value of this information is readily apparent to the quality control engineer who is attempting to produce the highest quality product at the lowest cost.

Conclusions

1. Quality control data for multiple correlations should be collected automatically by means of instrumentation.

2. A multiple-correlation analysis on a digital computer can effectively condense a vast amount of production data into concise mathematical equations.

3. These equations provide a sound basis for making operating decisions, optimizing variables, and furthering automation.

REFERENCES

Bennett, Carl A. and Franklin, Norman L., "Statistical Analysis in Chemistry and the Chemical Industries" - John Wiley & Sons, Inc. (1954)

Brownless, K. A., "Industrial Experimentation", 4th American edition H. M. S. Stationery Office, London (1953)

Ezekiel, Mordecai, "Methods of Correlation Analysis", 2nd edition Chapman & Hall, London (1941)

Greaves, M. J. "Statistical Computations Increase Production" - IRON AND STEEL ENGINEER, January, 1959 - pp. 99-102

Hart, Donald E. "The Digital Computer and Its Application" - National Convention Transactions 1957, 11th Annual Convention of American Society for Quality Control pp. 327-338

GENERAL LINEAR CORRELATION WITH AN IBM 604 ELECTRONIC CALCULATING PUNCH AND PLOTTING OF THE DATA WITH AN ELECTRONIC ASSOCIATES DATA PLOTTER

H. T. Oatman
Supervisor, Technical Data Center
Allegheny Ludlum Steel Corporation

When the Technical Data Center was established a number of years ago as a staff function of the Corporation Technical Department, many of our first problems were to investigate relationships between processing variables and end product properties or among end product properties themselves. The data, most of which was technical, was available from our Control, Chemical, Testing and Research Laboratories and some from production and inspection reports.

Some of the data was already available to us punched on standard IBM cards and much of it was in the form of separate sheets of paper of varying legibility and accuracy. Obviously, much clerical work was required to shuffle and reshuffle, transcribe and retranscribe the data to arrive at the point where computation could begin. From the above and other considerations, a decision was made to equip the Technical Data Center with standard IBM card handling equipment. To this was added the IBM 604 Electronic Calculating Punch which, although designed primarily to perform accounting functions, could also function as a general purpose digital computer. To simplify the equipment installation since most of our work would involve only statistical quality control and statistical analyses of data, it was further decided to restrict the equipment to function only in the numeric system rather than the alpha numeric system. The equipment of the Center was as follows:

IBM	026	Numeric Key Punch
IBM	056	Numeric Verifier
IBM	101	Statistical Machine (which could also function as a card sorter)
IBM	514	Reproducing Punch
IBM	077	Numeric Collator
IBM	419	Accounting Machine
IBM	604-521	Electronic Calculating Punch

One other piece of equipment was added primarily for use in our simple linear correlation work. This was the Data Plotter, made by Electronic Associates. Very early in my work I found, and in my humble opinion, it is imperative to plot the data before the correlation analysis is attempted. Our data for a given problem could vary from as few as 25 points to over 2500 points. We soon found it expedient to routinely use the Data Plotter for all studies involving more than 100 points. The value of the plot of the data is manyfold. Examination of the plotted points alone often indicated that a straight line would not fit the data. (See Figure 1). Plotting also quickly highlighted gross errors in the data contained in IBM cards. It was found not infrequently that our verifier-operator verified the mistakes of the key punch operator since neither could be certain of what was in the script data. The plots also helped us to interpret the results obtained particularly in respect to the possible effects of extreme values and clusters of data. (See Figure 2). To Management we were able to present many more graphical expositions of data than they had ever seen before. Since

there were no practical limits for volumes of data, generally each plot showed all of the pertinent data generated over longer time periods. Evident stratifications within the plots and nonlinearity were obvious. These and magnitude of the standard error shown for regression analyses (See Figure 3) proved thought-provoking material for Management. Unexpected relationships were demonstrated.

A brief description of the peculiarities of the particular equipment will be necessary since the equipment requirements always determine how the problem is handled.

For the Data Plotter to plot directly from IBM cards (we can and do use manual entry from a key board for small problems), the values of X and Y must be on the same IBM card. X and Y may be independently positive or negative in value and are limited to a maximum of four digits each. Negative values are sensed by the Data Plotter as a zone punch (11), in our case located over the high order or left most digit of the X or Y field of the card. Our Data Plotter is also equipped with symbol head enabling the Data Plotter to choose any one of twelve symbols for any point plotted. To use this feature an additional column is punched elsewhere in the card. An example of this is shown where a Maltese Cross may be chosen to indicate a longitudinal tensile value, and a circle a transverse tensile property when it is desired to show both on the same plot (See Figure 4). It is possible to use all symbols on the same plot.

Utilizing the 13-digit capacity of the accumulator unit requires that X and Y can have a maximum of 5 digits each and in this case the number of pairs cannot exceed 999. If X or Y is less than 5 digits, the number of pairs can be 9999. The sums of squares (X or Y) and also the cross products (XY) may not exceed 13 digits. In our program the sum of X or Y may not exceed 8 digits. The program to extract the square root necessary to give sigma will accept an 11-digit radical and give a square root of 8 digits. The above thus becomes the limiting condition for our general solution. To perform the computations when the maximum capacity is to be used requires that the 604 be equipped with a program repeat feature which doubles the time available to complete the calculations. We also require three sets of boards permanently wired. The first board will give us a card count (N), sums of the variables, sums of cross products, sums of the squares of the variables and sigma squared. The second board is a square root subroutine and gives sigma. The third board is used to compute the linear correlation coefficient r, the values of a and b.

Let us now begin with the incoming data. If the data comes to us on IBM cards, two possibilities exist.

1. Both X and Y are on the same card in which case negative values are already properly indicated. Symbol codes for use by the Data Plotter may have to be added by the key punch. These cards are then sorted for missing data in either X or Y or both. It is routine to indicate missing data with a preselected zone punch in one column of the field - this facilitates sorting. We now have a deck of cards containing X and Y and identity information for check purposes and properly coded. They are ready for the Data Plotter.

2. X and Y may be contained on separate cards; however, both cards contain common identity information. The X cards constitute one deck of cards and the Y cards the other deck. The IBM 101 functioning as a sorter or a sorter itself arranges the decks in sequence (heat and coil sequence, time sequence, etc.). The cards may also be checked for missing data. The two decks are next matched and merged using the IBM 077 Collator. In the merged deck, the card containing X is followed by the card containing the associated Y value. The IBM 514 Reproducing Punch generates a new third deck which now contains the values of X and Y and the common identity information and proper zone coding for negative values and symbol selection if any. The new deck is checked for missing data and then is ready for 604 processing.

If the raw data is in script form, we key punch an IBM card containing X and Y and common identity. These are verified and are ready for the Data Plotter. It may be more convenient to punch separate cards for X and Y, each card containing common identity when X and Y appear on separate sheets of paper, thus requiring the operator to scan and handle stacks of paper slowing her down no end and reasonably increasing the chances of errors. After verification of the cards, they are processed exactly like the previously described situation where X and Y entered the system as prepunched cards on separate decks. The result is as before - a single deck of cards containing the values of X and Y and identity ready for the Data Plotter.

Plotting of the data is now accomplished. The decks of cards containing both X and Y, identity, code, and zone punches are placed in the IBM 514 which now serves as a card reader wired to and controlled by the Data Plotter. The card is read. The Data Plotter finds X and Y on the plotting surface, the symbol is selected if this is required and the point plotted. The next card is called for and the process is repeated. Movement of the plotting head requires time so we find that we can plot about 35-45 points per minute, faster if the cards have been preordered so that X and Y are relatively of the same magnitude, slower if randomly ordered. Plotting proceeds automatically until the 514 is empty of cards.

At this point a word is required about setting up the Data Plotter. The X and Y axes may be independently scaled by a continuous factor up to 10 to control the size of the plot, and these factors are preset before plotting. The plotting surface, in our case, is approximately 30-1/4 inches by 30-1/4 inches (750 mm). We generally prefer to use standard size (8-1/2 inches by 11 inches, 11 inches by 17 inches) metric graph paper. The plotting surface itself is covered by standard 30 inch wide metric graph paper fed to it from a roll. Other sizes may be overlaid and secured in position by drafting tape if so desired. Correlation data is plotted in only one quadrant although the Data Plotter is capable of plotting in any single or any combination of the four quadrants. There is one restriction, however, which reduces the flexibility of the size of the scaling factors. X = 0, Y = 0 or the origin in any case must be on the plotting surface but not necessarily on the particular plot itself.

Normally, when plotting of the data is completed and visually examined for errors and nonlinearity, the deck of cards is transferred to the 604-521 for computation. However, when we are in an extreme hurry, we may have previously duplicated or reproduced the deck of cards being

fed to the Data Plotter. Such being the case, we can compute and plot at the same time, always hoping that there are no errors. Should the plot show errors, these, of course, have to be identified, the cards corrected and the computations rerun.

Normally, for computation the deck of cards is removed from the 514, a blank IBM card containing an X punch in column 80 is placed in back of the deck. This card will contain the results of the computations as will be evident as we proceed. A block diagram is given in the appendix, and I will only summarize it briefly here. Since our program is for a generalized problem of a maximum practical digital capacity, we have chosen to operate on X, Y, and XY separately.

The first run through the 604-521 gives us a card count (N), sum of X and sum of X^2 all punched into the blank card at the back of the deck suitably identified. This we call the "X" card. The "X" card is replaced with another blank card X punched in column 80 and the read-in wires changed on the board to pick up Y. The run is repeated generating a "Y" card which is in turn replaced by a third blank card and the run repeated (suitable change in read-in columns) to give us the "XY" or cross products card. The process is repeated the fourth time developing another IBM card containing the sum for X + Y and the sum of $(X + Y)^2$ designated as the check sums card. At this point our original deck of cards which may be as many as 999 or 9999 cards has been reduced to four IBM cards. Subsequent processing involves only these four cards (unless there are errors).

The four cards are now transferred to the 419 accounting machine for purposes of checking the calculations. The first pass through the 419 checks the sum of X plus the sum of Y against the sum of X + Y. A second run of the four cards provides us with the sum of $(X + Y)^2$ against the sum of X^2 plus twice the sum of XY + the sum of Y^2. When the check is validated, the fourth or check sum card is discarded reducing the working cards to three. This step as well as the next could be readily done on another pass through the 604, but this would require an additional board. Next, it is necessary to pick up the sum of Y from the "Y" card and reproduce this into the "XY" card and also introduce an X punch in column 77 for later processing. This may be done by the key punch or more regularly by the 514.

The three cards are now returned to the 604-521 and passed through for computation of $\bar{X}$ and $\bar{Y}$ and punching of these statistics in their appropriate cards. The XY (cross products) card is ignored in this pass. The next pass through computer punches in the appropriate X and Y cards σ^2 . The covariance of XY is computed for and punched in the cross products card as it passes through. Board 1 is replaced by Board 2 which contains the program for the square root subroutine. The next pass of the X and Y cards, each now containing their respective σ^2 , results in the computation and punching of the sigmas.

Now, as the result of reducing our original deck of cards in four passes through the 604-521 in four passes to four IBM cards, an independent check on auxiliary equipment and two more passes through using Board 1 and one more pass where Board 1 has been replaced with Board 2, we have as an end product three cards containing the necessary quantities required for the computation of r, b and a as well as auxiliary

quantities not subsequently used.

The 521 now selects from three cards the desired values and punches them into a single card which now contains sigma X, sigma Y, $\bar{X}$, $\bar{Y}$, and the covariance of (XY).

Board 2 is now replaced with Board 3, and the single card is passed through resulting in the computation and punching into that card the values of r and b. The next and last pass similarly develops and punches the value of a.

The card is now read out and printed by the 419. With these values, we write the equation of the straight line in the intercept form, solve it for two points and transfer to the plot of the data, adding the standard error lines if required and are done.

We are frequently asked how much time is required to do all this. The most useful answer is, of course, related to the data volume. Since the 604-521 unit operates at 50 cards per minute (using program repeat), the card volume divided by 100 and this value multiplied by 4 for the first 4 passes plus a minute or two for card handling accounts for most of the time required for a large problem. Beyond the stage where the original deck has been reduced to four IBM cards, approximately twenty minutes are required for completion of the solution. The Data Plotter operates roughly at a speed of 40 cards per minute. The plotting time can be readily determined. As mentioned earlier, if the problem is large, it may be expedient to use duplicate decks of cards. In this case the plotting time and computing time are not additive.

This paper has been presented that it might prove useful to listeners and readers who may have access to an IBM 604-521 unit in their accounting departments. Admittedly, these may not be full capacity units. The cost of building the 604-521 unit up to full capacity 60 program steps, the addition of the program repeat and punch delay features are not prohibitively expensive. The Data Plotter is a very useful but expensive accessory unit. However, we were fortunate that very early in its use we were asked to demonstrate graphically a situation which, in itself, more than justified the cost of the equipment. Since then, we have been riding the gravy train so to speak. The data processing system described is not necessarily the best or only one. Actually, we have dispensed with the 604-521 unit replacing it with an internally programmed paper tape computer. However, the change was not attributable to the problem of a general program for the determination of the correlation statistics.

For purposes of unifying this portion of the panel discussion, I have been asked to submit to the equipment manufacturers five questions. They are:

To Mr. Mario V. Morreale, representing IBM

1. No attempt was made to optimize the above program. Any suggestions of this nature would prove of help to anyone wishing to use this method.

2. Although our program for Board 1 is only 50 steps, we find that we have to use program repeat since apparently the internal timing of the machine functions require this. Any comments on this situation would be helpful.

To Mr. James N. Smiley, representing Electronic Associates

1. I believe the Data Plotter we have was the eleventh such instrument of this particular model. Earlier mention was made of the fact that the origin of the plot must be located on the plotting board. Why is this so and do the later models have this requirement also?

2. For automatic operation, we read IBM cards. Can the Data Plotter be equipped to read data from paper and magnetic tape perhaps with auxiliary equipment under the selection of the operator? I ask this question since data is coming to us in two of the three forms already and the third is not far off.

In closing, I wish to express my deep appreciation of the work of my associates, Miss Betty Caugherty and Mr. Norbert Lauer, who were chiefly responsible for developing the data processing system and programing the IBM 604 and Data Plotter.

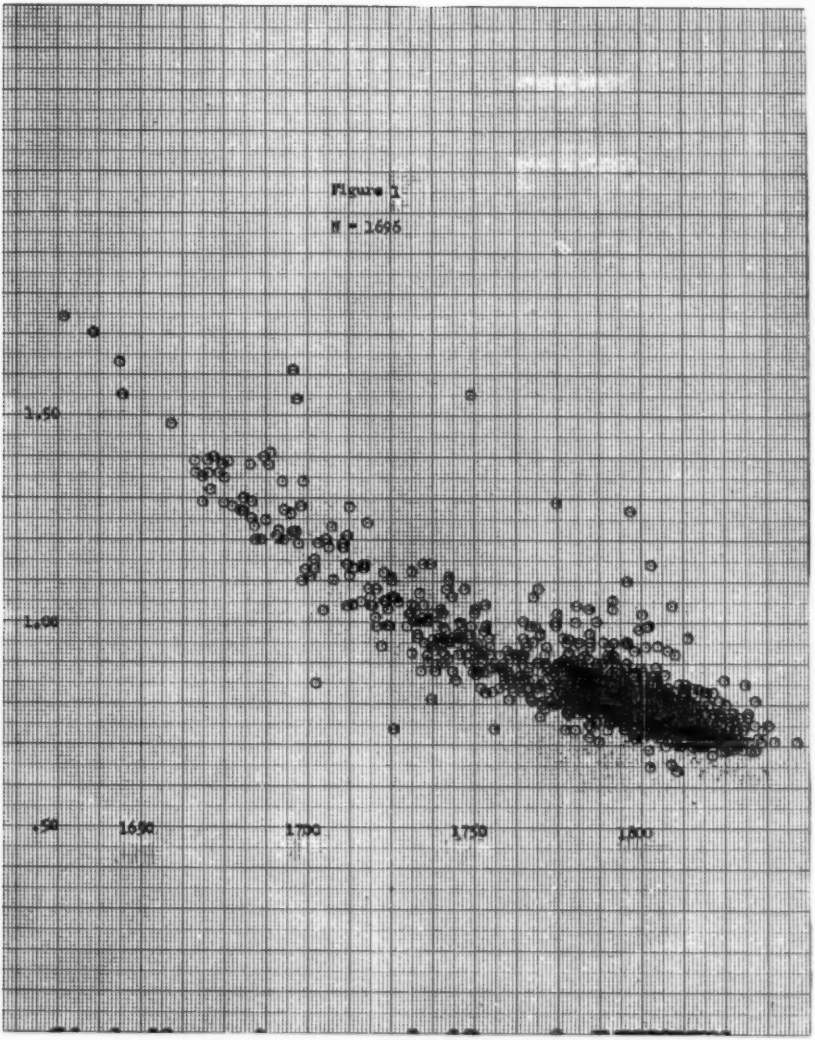


Figure 1
N = 1695
1.50
1.00
.50
1650
1700
1750
1800

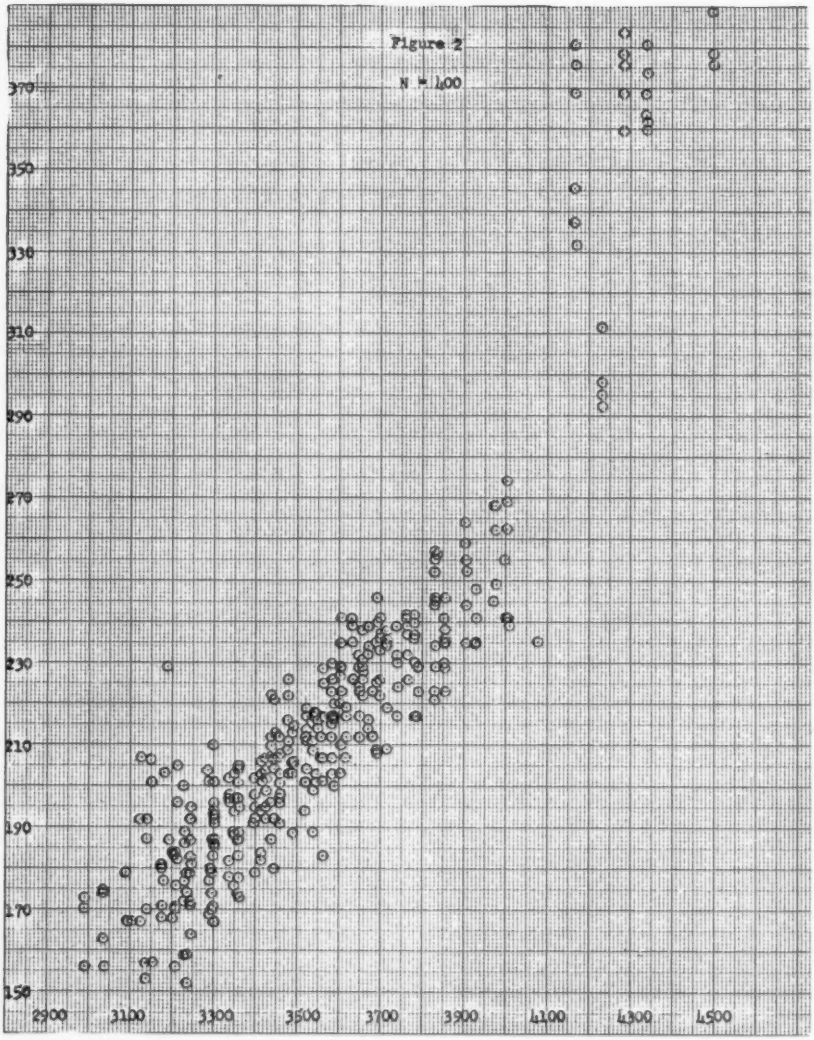


Figure 2
N = 400
370
350
330
310
290
270
250
230
210
190
170
150
2900
3100
3300
3500
3700
3900
4100
4300
4500

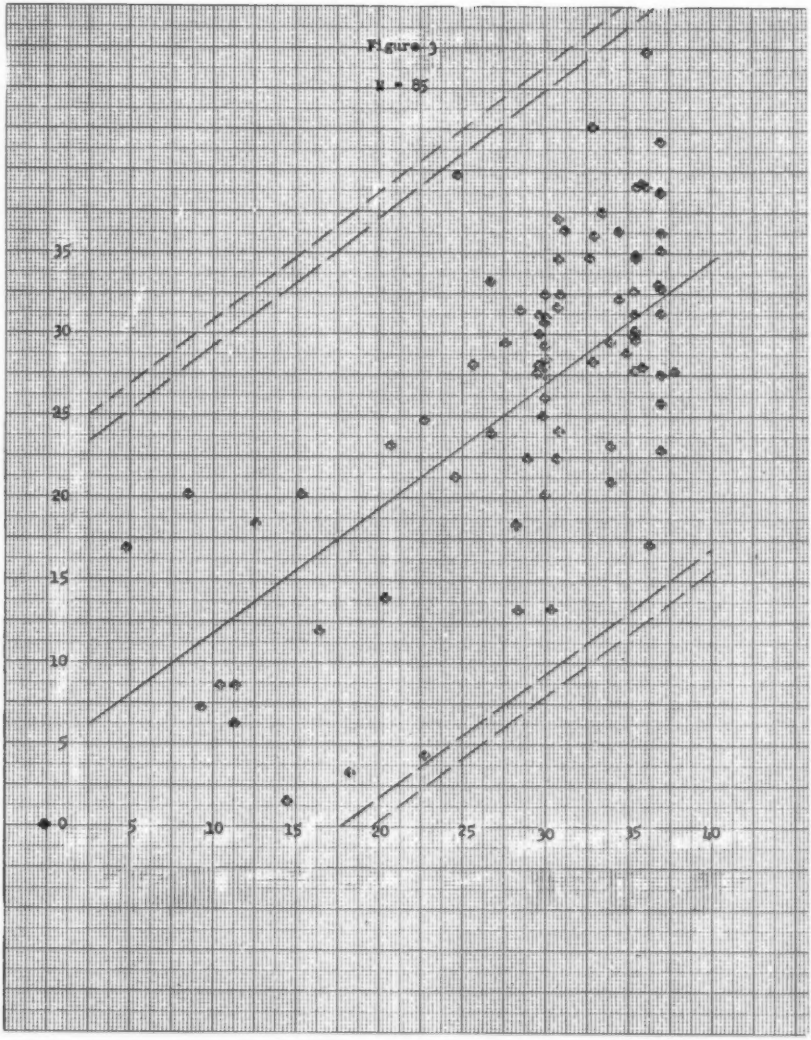


Figure 3
N = 85
35
30
25
20
15
10
5
0
5
10
15
20
25
30
35
40

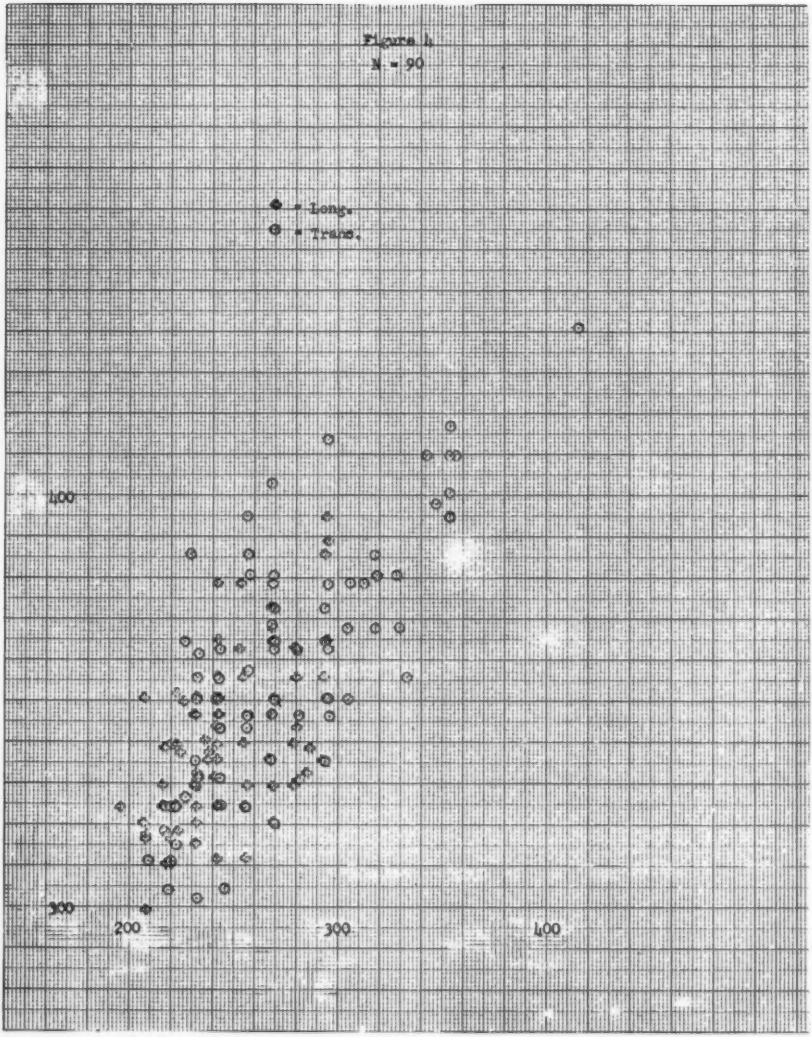
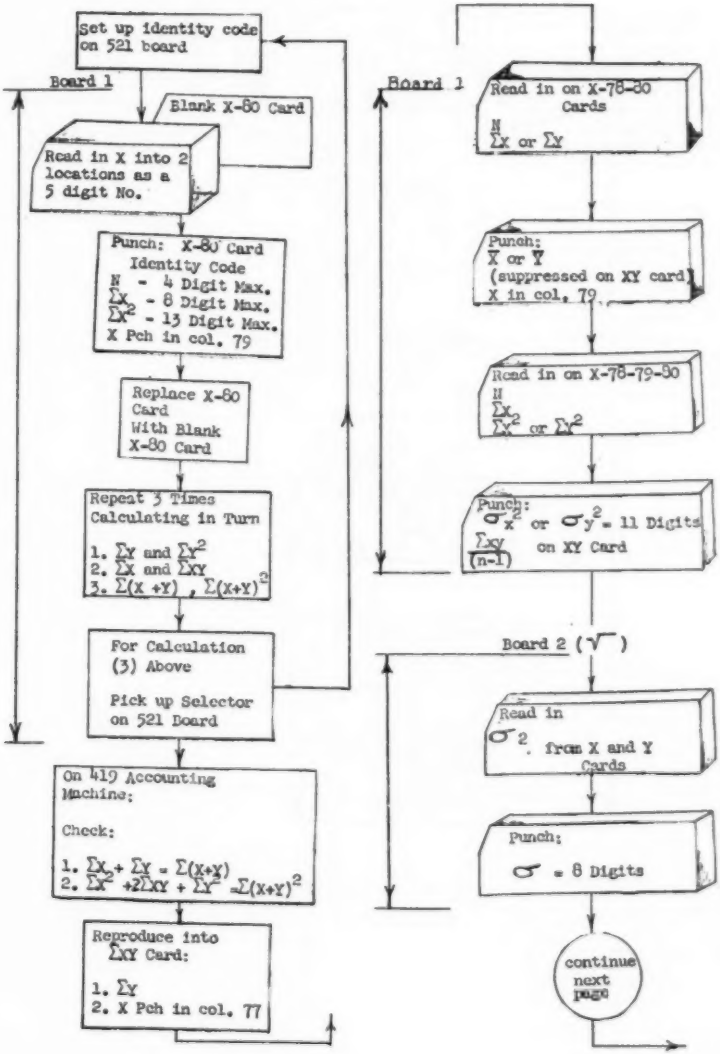
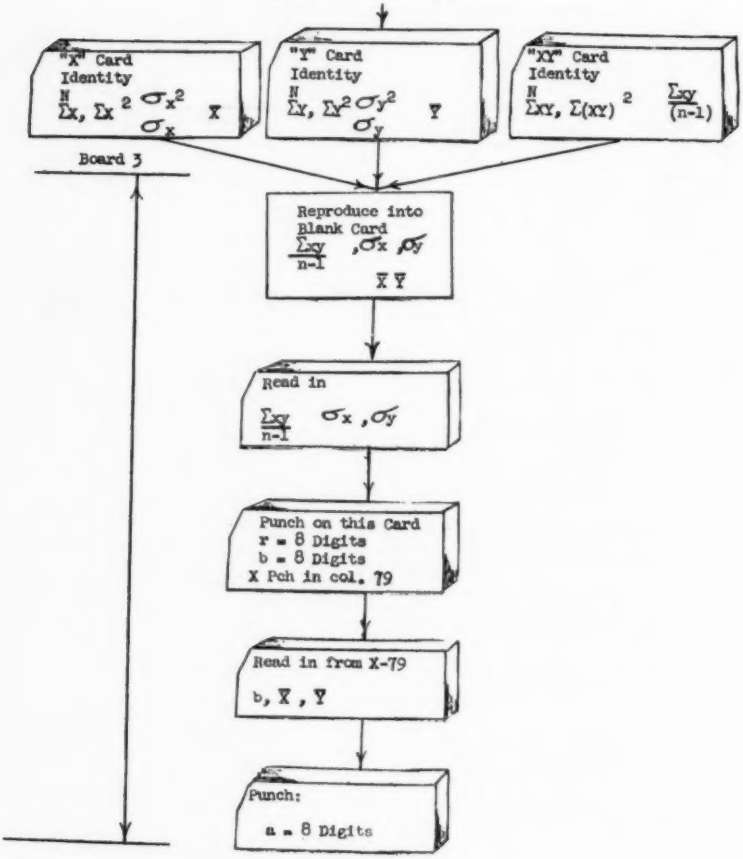


Figure 4
N = 90
= Long.
= Trans.
400
300
200
300
400

Appendix I
Block Diagram IBM-604 For
General Solution Of Linear Correlation Model



Appendix I cont'd (2)



Appendix II
Equations Used For General Solution Of
Linear Correlation Model

Calculate N = Card Count

$$\bar{X} = \frac{\Sigma X}{N}$$

$$\bar{Y} = \frac{\Sigma Y}{N}$$

$$\sigma_x = \sqrt{\frac{N\Sigma X^2 - (\Sigma X)^2}{N(N-1)}}$$

$$\sigma_y = \sqrt{\frac{N\Sigma Y^2 - (\Sigma Y)^2}{N(N-1)}}$$

$$r_{xy} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{N(N-1)\sigma_x\sigma_y}$$

$$b = \frac{r\sigma_y}{\sigma_x}$$

$$a = \bar{Y} - b\bar{X}$$

THE FIRST FOUR MOMENTS OF A FREQUENCY DISTRIBUTION AND A PLOT OF THE DISTRIBUTION

Philip Hermann
Jones & Laughlin Steel Corporation

INTRODUCTION:

As users of modern data processing equipment in implementing mathematical statistical techniques we have operated on a do-it-yourself basis. That is, our statisticians have become proficient in the basic programming and operation of the computers (IBM 650) and associated tabulating equipment (sorters, printers, interpreters, and punches). Our experience as a small group (5) has confirmed this policy as being successful in greatly increasing the effectiveness of the statistician. His familiarity with the capabilities of the data processing equipment has led to more thorough and complete analyses than would otherwise be possible.

The "solutions" to the problem presented to us are a product of this policy. We have made use of machine language programming, various interpretive routines, assembly programs and, more recently, appropriate compilers in writing programs for the IBM 650 computer. The particular programs cited below are not necessarily final solutions; as more efficient program methods are developed and the need for more efficient utilization of computer capacity arises, they will be modified.

PROBLEM:

From our point of view the three different card volumes (100, 1000, and 20,000) pose the same problem. To obtain the first four moments of a frequency distribution involves considerable, repetitive numerical operations. Thus, for purposes of speed and accuracy, a machine computation is desirable.

The plotting of the frequency distribution doesn't really present a problem at 100 sets of observations, but the time for tallying and plotting becomes significant, certainly, at 1000 or more sets of observations.

To date, we have resolved the computational problem by the use of interpretive routines in programming an IBM 650 computer (standard components) to perform these calculations. The frequency tally and distribution has been programmed on the IBM 650 computer and 407 accounting machine.

SOLUTIONS:

First Four Moments of a Frequency Distribution

In particular, to calculate the first four moments of a frequency distribution we have used the modification of Technical Newsletter #11 of IBM's Applied Science Division (A Complete Floating-Decimal Interpretive System for the IBM 650 Magnetic Drum Calculator, by V. M. Wolontis, Bell Telephone Laboratories, Inc., Murray Hill, New Jersey) developed by the Computing Center in collaboration with the Statistical Laboratory both of Case Institute of Technology (A Statistical Interpretive System for the IBM 650 Magnetic Drum Calculator, by G. E. Haynam, Cast Institute of Technology.) These interpretive routines take statistical data fixed decimal form.

If the raw data is already on punched cards, this will involve a conversion operation to the statistical input card format. If the data is not punched, it can be put directly into the proper card format using an IBM punch machine (026). This format, aside from certain indicative information for purposes of control as well as identification, will contain one six digit fixed point observation per card. This conversion, of course, will depend upon the particular tabulating installation with respect to availability of equipment and economic usage of same. One straightforward approach is to hand punch the required input deck, copying from a listing of the original data. Another method is to use a reproducing punch to read the original deck and through the appropriate wiring pattern to punch out a revised deck in the proper format. A third method is one of wiring the 533 panel of the 650 to accept the arbitrary format; this is usually the most economic method unless there are a series of such problems to be run and it is essential to conserve computer time. If a mathematical transformation of the data is desired, then it would be more economical to program the 650 along with the appropriate wiring of the 533 panel to read the original data deck and perform the desired mathematical transformations and punch out the revised data deck. (The 650 can act as a reproducing punch, but this is not at all economical, except in case of breakdown of the reproducing punch (514), and where time has at least a dollar value equal to the hourly rental of the 650).

Once the statistical input deck has been made it is a simple matter to use the Statistical Interpretive Routine with its wired control panel and a simple program utilizing the statistical commands for each of the four moments.

FREQUENCY DISTRIBUTIONS:

Relative and Cumulative (Programs written by Arthur Evans, Jr., formerly of the Applied Mathematics Section of J&L Steel Corporation, now of the Graduate School of Carnegie Institute of Technology).

The frequency distribution program utilizes the IBM 650 computer to read a deck of cards with data punched on them--in any fixed decimal format--to generate a card deck for use on the IBM 407 Accounting Machine. The output from the 407 is a point graph on which the abscissa is class interval upper end points and the ordinate is frequency in the class interval.

The program is capable of producing either a frequency distribution or a cumulative frequency distribution. To operate, the program needs to know the range of the data and the number of class intervals desired.

The output deck consists, essentially, of one graph card for each class interval defined, with some extra cards generated by the program. The scale of the graph will always be adjusted so that the largest ordinate is at least half the graph, and is printed at the bottom of the page. The first line of printing below the printed scale gives the number of observations in the input data deck; this card is generated by the program.

At the left of the graph, the tabulator gives the number of observations or the relative frequency of the observations in the class

interval whose upper limit is indicated (the upper limit is inclusive). The graph may be double or single spaced. The class intervals can recognize up to 9 digit numbers if so desired. The graph may be plotted with any character that the 407 is capable of printing, except the zero, minus, or 12 punch.

Time Series (by Arthur Evans, Jr.)

A companion program "Time Series" can be used, along with the Frequency Distribution board for the 407, to plot point graphs. TS reads cards one at a time. From each card it expects to find an X value (time axis) and a Y (observed variable) value. The X value which may be up to 9 digits, is merely reproduced for printing in the output. The Y value, which may be 5 digits, is plotted. That is, it is scaled between zero and 100 (according to limits previously given to the program), and a symbol is printed on the graph at the appropriate location. As in Frequency Distribution, a scale is printed at the bottom of the graph, followed by an observation count and a final comment card. Also, comment cards as the user desires may precede the graph.

With regard to the X values on the input cards, the program can operate in three different modes: PRINT, CHECK and ADD. The PRINT Mode was described above; in this mode the program merely reproduces for printing in the output the X value on the card. In many cases each X value is expected to differ from the preceding value by a constant, it being desired to skip a space on the graph if an X is missing. This can be done by operating in the CHECK mode. In this mode, the program is supplied with the first X to be read (X_0) and the desired interval between X values (D_x). If between one and ten values are missing, the program will supply the appropriate number of cards which will print X values but will not cause the plotting of a graph point. (If more than ten values are missing, or if the new X differs from the last X by other than a multiple of D_x, there will be an ALARM stop.)

Finally, in the ADD mode, the program will generate X values itself, and will not expect to find them on the data cards. Here again the program must be supplied X_0 and D_x.

BIvariate Distribution (Program written by D. C. McCune of the Applied Mathematics Section of J&L Steel Corporation.)

An additional program has been written to generate an x-y frequency distribution. The operator specifies the lower limit of both X and Y, the cell sizes (class intervals) and the number of X and Y cells. A maximum of 40 X or Y cells may be used. Data is fixed decimal up to ten digits in length with a maximum cell frequency of 10^{10}. Input data format is controlled by wiring the 533 panel. Output is one card for each non-zero frequency in the XY table giving the X and Y upper cell limits, the cell frequency, and the total frequency. The program operates at read speed (200 cards per minute).

REMARKS:

To facilitate the general use of these programs, appropriately wired 533 and 407 boards are maintained. For the Wolontis interpretive routine one 533 control panel is permanently wired; for the Frequency and Time Series programs a 533 panel and a 407 panel are permanently wired; the Bivariate distribution program uses any 533 panel which will

allow 2 word input and 4 word output; and the Wolontis and Bivariate programs use a general purpose 80 column wired 407 panel. An Interpreter can be used instead where tabulating on a general purpose 80 column wired 407 panel.

These programs have been found to be very effective in reducing the amount of clerical effort required in processing statistical data. The first four moments have usually been part of an overall program; in our case they have been part of a general program for evaluating frequency distributions on the basis of Pearson's System of Frequency Distributions. The first two moments are frequently incorporated in various statistical routines.

Detailed instructions and program decks for the programs described above (or similar ones) will be available from the program library of IBM.

A FIVE VARIABLE MULTIPLE CORRELATION STUDY ON UNIVAC I

Donald S. Macoy
Staff Supervisor, Computer Programming, Gary Steel Works
United States Steel Corporation

Multiple correlation problems are not "solved" on computers in the same way that we consider engineering or mathematical problems solvable. In theory at least, in multiple correlation, masses of data are merely organized so that it becomes possible for limited human faculties to handle the analysis in order that the various elements may be distinguished and their significance appreciated. The solution through multiple correlation (if indeed a solution is possible by this technique) results only from the analysis of machine output by the trained statistician. However, all too frequently we process mountains of data of dubious quality on expensive equipment to arrive at mouse-sized conclusions. And too, the recurrance of problems previously solved by correlation studies leads one to wonder whether the original solution resulted not from the statistical analysis, but rather from the improvement one would normally expect when the attention of experts in quality control industrial engineering, research and operations is focused upon a particular problem area.

Any consideration of organizing data for statistical analysis on a machine system such as Univac I must recognize that it costs between $100 and $300 per hour to operate and, while it is quite easy to look upon such analysis as "applied research" to which the application of a definite dollar value is difficult, infeasible or perhaps even unconscionable, we do owe our companies the obligation of getting the job done at the least possible expense consistent with timeliness and accuracy.

If this approach is accepted as reasonable, then it follows that Univac I would not be selected as the vehicle for a five variable correlation study, the card volume (or number of observations) being an immaterial factor. At United States Steel for example, a five variable study would be accepted as a Univac I application only if time were available which could not be utilized more advantageously, or if IBM 650 time were not available.

A second but equally important consideration to the person responsible for efficient equipment utilization, and one which also is a function of cost, is an assurance of the applicability of the data to the proposed analytical technique. It is not necessary to dwell upon this matter at a Quality Control meeting except to indicate that one would at least want to look at the raw data as regards homogeneity, linearity, adequacy of sample, independence of "independent variables", and perhaps in some instances the need for a large number of observations. The point is simply that the machine aspects of correlation analysis have been routinized so that anyone can perform the calculations. Quite often those who are neither familiar with the technique nor appreciate its pitfalls are apt to say when other analysis fails; "Let's try multiple correlation" without really considering the applicability of the data or the need for careful training and experience in interpretation of results.

Assuming, however, the acceptance of a five variable multiple corre-

lation as a Univac I application, there are two approaches to solution both of which are quite similar since both are based upon matrix manipulations.

(1) The matrix Math Instruction Tape
(2) A compiler system

The matrix math instruction tape is a set of computer programs each of which performs a particular matrix operation, i.e. transposition, multiplication, inversion, etc. A compiler is a sort of "super" computer program which acts upon a library of manipulative routines (usually an adaptation matrix math tape) selecting the desired operations under control of a "specifications" tape. The compiler assembles these operations onto a specific instruction tape in such a way that computation is continuous. For example, if one wanted to calculate the sums of squares and cross products of a matrix using the compiler system he would specify first the transposition routine and second multiplication. The compiler would search the library instruction tape for these routines, assign servo allocation and provide an instruction tape which when applied to data would perform the desired functions.

Of course the same results could be gained by using the matrix math tape directly without recourse to the overriding compiler. A small savings in computer compiling time would result which, however, would probably be more than compensated by increased handling time. There are several distinct advantages to the compiler system which will lead us to favor it in most instances.

(1) There is much less need for operator intervention in the computer processing.
(2) There is little, if any, need for services of the mathematician or statistician during calculation when the compiler system is used. The mathematician or statistician is intimately concerned with preparation of the specifications tape.
(3) Tape changing is minimized.
(4) Since less human effort is required there is less chance for error.

In a problem involving many variables and/or many observations, the advantages of the compiler system over the matrix math instruction tape system become less significant. However, where the number of variables or observations is small as in today's problem the advantage of the compiler system is clear. In general, the card volume (or observation volume) is an insignificant matter in determining processing technique, as long as it meets the conditions of necessity and sufficiency.

There remains then only to decide what is meant by multiple correlation and how the computer calculation shall proceed. Though the term multiple correlation means essentially the same thing to all statisticians, each has his own particular requirements regarding output from a computer correlation study. Of course, an instruction tape could be compiled which would fulfill the needs of all staticians - exceeding the requirements of some. This, however, while convenient is expensive in terms of computer time. At Gary Steel Works of United States Steel we have several program tapes tailored to the needs of particular statisticians who use our Univac. By this means we give each person

the information he wants in the sequence in which he wants it, but no computer time is wasted. Exhibit A is a sample of the "specifications" tape for multiple correlation using the Franklin Institute compiler. Other compiler systems are available; for example one has been developed by National Tube Division of United States Steel. Supplemental information could be made available by simply adding necessary matrix manipulation routines at appropriate points in the specifications.

Concerning the question of whether calculation should be interrupted at intermediate check points for examination of results, there is divided opinion among statisticians in my company. One of our most successful metallurgists examines his data very carefully while it is in card form and then proceeds to run his regression studies all the way estimating equations without stopping. In general, I believe this to be an unsafe practice, but I cannot argue with his success. I prefer to review the means for reasonableness before calculating sums of squares and cross products which is the longest computation in the series. Also I like to examine the simple correlation coefficients and the coefficient of determination as a clue to whether further calculation is warranted. The computer time spent in interrupting the machine process has proved worthwhile.

Univac I is a good machine for large scale correlation studies, those with more than, say, 10 variables. If I were asked to make a single recommendation to the manufacturer which would prove most beneficial to us in statistical studies, I would suggest that a multiplication routine be written which would take into account the symetrical nature of matrices used in multiple correlation. Computer time could be cut by at least 30%. Finally, since United States Steel does not usually calculate five variable regression studies on Univac, I am unable to provide empirical time estimates. The manufacturer does have estimating equations which may be valid for a problem of this size.

MULT COR #2

EXHIBIT A

OP.#	JMM FEB 1959	
1	MTMEAN000264 01LISTX-M	"Centering" of X (independent variable) matrix X → X_c (Means on SCP or Servo 6 by SCK option)
2	MTMEAN000379 01LISTY-M	"Centering" of Y (dependent variable) vector Y → Y_c (Means on SCP or Servo 7 by SCK option)
3	IN EDT000405 00TAPEX-M	Edit X matrix
4	TRANS 000584 00TAPEX-M T	Transpose X matrix
5	0+MULT00045- 00TAPEC(XX)	$X_c^t \cdot X_c \rightarrow C_{xx}$ ("Cross products matrix")
6	DIUFIL000-02 00TAPEC(XX)*	Diagonal fill before inversion of C_{xx} (Tape C_{xx} retained for output edit.)
7	INVERT00020- 01TAPECXX-1 -00000000009	Invert C_{xx} (Tape C_{xx}^{-1} retained for output edit.)
8	IN EDT000902 00TAPEY-M	Edit Y matrix
9	0+MULT000429 00TAPEC(XY)	$X_c^t \cdot Y \rightarrow C_{xy}$ ("Cross products vector")
10	0+MULT000-94 00TAPEB	$C_{xx}^{-1} \cdot C_{xy} \rightarrow B$, regression coefficients Coefficient vector retained for output edit)
11	0+MULT00054- 00TAPEX*B	$X_c B$ = estimated Y vector; Y'
12	SUBT 0002-3 00TAPEY-Y'	$Y - Y' =$ residuals
13	NORM 000300 00TAPE000000	Divide by " $\sqrt{N-1}$ " to get S_y, standard error of estimate
14	TRANS 000903 00TAPECXY T	Transpose "cross products vector"
15	TRANS 000205 00TAPEY-M T	Transpose Y_c vector
16	0+MULT00052- 00TAPEYT*Y	$Y_c^t \; Y_c \rightarrow Y$ sum squares
17	NORM 000-06 00TAPE000000	
18	DIUFIL000-05 01TAPEYT*Y * 000000000009	Invert $Y_c^t \; Y_c$
19	INVERT00050- 01TAPEYT*Y-1 -00000000009	
20	0+MULT0003-5 00TAPERXY T	Generate multiple correlation coefficients of Y on X $R_{XY}^t = C_{xY}^t (Y_c^t Y_c)^{-1}$, and copy for output edit.
21	0+MULT000549 00TAPE*R*	
22	NORM 000900 00TAPE000000	$R_{xY}^t * B = R$, the multiple correlation coefficient
23	ZZZZZZ0000T2 000000000000 000000000000 000000000000 000000000000	Output edit C(xx), $C\,(XX)^{-1}$, B, $Y-Y'$, $Y^t \cdot Y$, $R\;(XY)^t$, R^2

A FIVE VARIABLE MULTIPLE CORRELATION STUDY ON UNIVAC I

Arthur J. Flynn
Remington Rand UNIVAC

The reputation and continuing success of the Univac computers depends primarily on their being used profitably by our customers. Remington Rand UNIVAC values highly its skillful and judicious computer users --such as the United States Steel Corporation. Mr. Macoy, of U.S.Steel's Gary Works, has disucssed several topics pertinent to the appropriate use of their Univac I. I would like to comment on one or two of these.

It is significant that applied statistics has held a prominent position in the developments of computer applications. Those statistical techniques which are designed for the analysis of large quantities of data or of many relationships--involving, as they do, a great number of calculations--provide excellent utilization of the data handling capacities and the computing speeds of the large scale computers. Also, the present stature of statistical quality control demonstrates the practicality of using the powerful techniques of modern statistical analysis.

In order to facilitate and economize the transition from the statistician's problem to the computer's operation, Remington Rand and Univac users have developed Univac Library Routines (e.g., the Statistical Moments routine, the Matrix Operations routine, the Random Number Generators) and program-generating compilers--such as the Matrix Mathematics Compiler. These comprehensive compilers can also provide validity-checking of original data and of intermediate and final results. However, these will not compensate entirely for inappropriate analysis or for erroneous or unsuitable data. Accordingly, the statistician's knowledge and judgment must play a major role in the use of the computer. The approach taken at U. S. Steel of tailoring computer programs to the individual needs and preferences of the different statisticians supplying computer applications is particularly valuable in this regard.

Mr. Macoy has pointed out that some proposed applications--such as an isolated, one-time, five variable multiple correlation problem--may not be of an appropriate size for typical, routine Univac I operation. When we consider preparation of input data, compilation of program, loading of program and data, and operational set-up of computer and peripheral equipment--all to accomplish perhaps five minutes of computer calculation, such applications cannot compete favorably with more extensive ones. Also, any serious appraisal of computer operating costs will include the time of programming and operating personnel, plus the machine time of all of the devices involved--not to mention power, maintenance, overhead, etc.

Those who also have available a small or medium-size computer have a nice alternative. For those who do not, there is another approach which may merit consideration. This is to accumulate a sufficient number of these small applications to comprise a reasonable computer run--programming them collectively by means of an extensive compiler. This compiler would either generate a program providing all typically desired measures, or would select all of the appropriate sub-routines for each application on the basis of minimal statement of requirements. Obviously computer time is not optimized--in the one case during computation, in the other

case during the compiler run. However, minimizing preparation and handling should more than compensate for this.

To avoid misconceptions, I feel that I must emphasize the following point. It is strictly the limited size of the five variable multiple correlation which may make it inappropriate for the computer. In general the multiplicity of variables in correlation, regression, variance analysis, or linear programming problems is a principal factor placing such problems among the most valuable and appropriate applications achieved on the large scale computer.

Computers are, of course, just new tools--very capable, quite expensive, and potentially of great value to us. It is most encouraging to have people in statistical quality control--well-versed as you are in the development, utilization, and evaluation of new tools and methods--contributing to the future of the computer field.

IBM DATA PROCESSING EQUIPMENT AS APPLIED TO QUALITY CONTROL

Mario V. Morreale
International Business Machines Corporation

Problem:

The first four moments of a frequency distribution and a plot of the distribution when 100, 1,000 and 20,000 cards are to be reduced.

Solution:

In answer to our user, I agree that the computation of the first four moments involves considerable, repetitive, numerical operations. Thus, for speed and accuracy a machine computation is of necessity.

As for the tallying and plotting of a frequency distribution, I would also recommend the use of a computer.

As indicated, your answer to the problem has been the use of interpretive routines to calculate your moments. The use of these interpretive routines is excellent as the first solution to the problem. If it is desired to reduce the time factor, it is my suggestion that a program be written utilizing direct machine language.

We have written such a program. In fact, as a matter of interest, the program is part of our exhibit.

The program operates in fixed decimal. The number of observations are "n" and the number of variables can be "k". The program operates on one variable at a time. It is necessary that the individual requesting the data reduction give the cell boundaries (maximum of 10) a "high" and "low" reading so that true mavericks will not be computed as part of the statistics.

The output of the program is at present a printed report. With modifications, a punched card output can be obtained if so desired. The report gives the following information: mean, variance, standard deviation "n", a frequency distribution; both the actual number and as a percent to the total number of observations (mavericks included). Further, it gives a tally of the frequency distribution and the cumulative distribution. If it is desired, anyone of the last two phases can be eliminated, i.e. the tally or cumulative distributions.

The time involved for complete outputs are approximately 8 minutes for 1,000 observations for any one variable and 2 1/2 hours for 20,000 observations.

ALPHA IBM TYPE 065 03099

N	MEAN	VAR	STDEV	MAVHI	MAVLO
100	99.28	182.04	13.49		

FREQ. DIST.

20 64	65 74	75 84	85 94	95 104	105 114	115 124	125 134	135 144	145 170
1	2	9	22	33	22	8	2	1	

FREQ. DIST. IN % TOT N

20 64	65 74	75 84	85 94	95 104	105 114	115 124	125 134	135 144	145 170
1.0	2.0	9.0	22.0	33.0	22.0	8.0	2.0	1.0	.0

TALLY IN % TOT N

20	64	1
65	74	11
75	84	111111111
85	94	1111111111111111111111
95	104	11111111111111111111111111111111111
105	114	1111111111111111111111
115	124	11111111
125	134	11
135	144	1
145	170	

```
%MORE THAN AND %LESS THAN
100   2    2    2                    1    1    1    1
 95
 90                   2         1
 85
 80
 75
 70                        2
 65
 60
 55
 50
 45
 40
 35                   1         2
 30
 25
 20
 15              1                   2
 10
  5   1    1                                2    2
                                                      2
     64   74   84   94   104  114  124  134  144  170
UPPER CELL LIMIT
```

Problem:

Simple Correlation Coefficient.

Solution:

In answer to Mr. Oatman as to any suggestions on how I might improve his application of an IBM 604-521 in the solution of the problem of finding a simple correlation, may I first say that he has utilized the equipment very effectively. I concur data should be verified and plotted prior to any processing through a computer.

To answer question No. 1, I cannot add to the optimization of your program since it appears that all details have been accounted for and you are using a 604 with all the special features.

Again, question No. 2 must be answered in the negative. Program repeat must be used to satisfy the timing of the machine functions.

My only suggestion is that the next level of computer be used; namely, our transistorized 608-535 calculator. The 535 operates at a speed of 155 cards per minute as opposed to the 521 which operates at 150 cards per minute. More program steps and storage is available. Non-sequential programming allows the program steps to be used in any sequence without loss of time for skipping unused steps. Storage consists of forty (40) nine-digit words which can be split into three and six-digit words with separate signs.

If it is desired to have even more expediency in the solution of a statistical problem and the budget permits, an IBM 650 would be more than ample in the solution of the basic and advanced statistics confronting Quality Control people in controlling production processes.

HOW STATISTICAL QUALITY CONTROL IS USED IN FOOD PROCESSING INDUSTRIES

Howard L. Stier
National Canners Association

Uniformity in the quality of a processed food product is dependent to an important degree upon effective quality control of the raw product. In the canning industry the production of such raw materials as fruits and vegetables is always subject to very close supervision by the canner in order to meet his standards of quality in the processed product. In some instances the canner finds it necessary to produce the raw material himself in order to obtain the quality and quantity required. When the fruit or vegetable is grown for the canner by farmers, the canner in many instances provides for the farmer many of the services of production that are essential if the quality standards are to be attained. Some of the most advanced and sophisticated techniques of the science of statistics have been used in the production and quality control of vegetables and fruits for processing. Indeed, many of the first applications of modern experimental designs and the use of the analysis of variance in the interpretation of experimental data were in this field. Despite this long history of the effective use of statistics in the control of raw product quality the application of the same techniques to the processing operation has been relatively slow. Ten years after the successful use of statistical quality control techniques by other manufacturers there were very few food processors who were using these procedures. Indeed, until recently only a few food processors seemed to have any knowledge of the techniques and their possible application to food processing. However, in the last five years there has been a marked acceleration in the interest and application of such techniques in food processing. This slowness of the food processing industry to use SQC techniques has been attributed to the high variability inherent in food processing (9), lab methods of analysis which provides answers many hours behind the process (14) and the inherent problems of organoleptic evaluation of quality. However, as Harrison (9) points out "whether such problems are regarded as an obstacle or a challenge will depend upon the morale of the industry."

To the author's knowledge the first conference devoted exclusively to the application of SQC in food processing was held in February, 1952 at the University of Maryland as a joint undertaking of the National Canners Association and the Department of Horticulture of Maryland. Evidence of the beginning interest in this application to food processing was clearly indicated by the geographic distribution of the people who attended that five-day conference. The conference was originally planned as a short course on SQC for food processors in the Middle Atlantic area. However, some

companies apparently felt that this new technique had sufficient potential value in food processing to be worth sending representatives of their company from Maine, Colorado, Illinois and even from the state of Washington to College Park, Maryland. The interest developed by food canners and freezers after this first session was sufficient to warrant repeating at a later year in March, 1953. In November, 1955 a similar five-day short course in the application of statistical quality control in the canning industry was conducted at the West Coast Laboratory of the NCA in Berkeley, California. A total of 45 persons from 26 canning companies attended this shortcourse. This period marks the beginning of the most rapid increase in the application of statistical quality control in the canning industry. Since that time the number of companies that have begun to use statistical quality control techniques has increased rapidly. In addition, the types of applications have expanded.

Today the techniques of statistical control are being used in food processing in the evaluation and control of raw product quality, quality of containers, equipment performance, processed product quality, container fill and weight control, labeling, quality of purchased supplies and processing efficiency. Although there are very few applications in the following fields, they would seem to offer possibilities in the further effective use of statistical quality control techniques: sanitation, safety (work injury control), control of damage in transit, production worker efficiency, cost control, delay allowance and causes, as well as in the fields of clerical operations and distribution and marketing.

In 1955 the Food Industries Committee, then a section of the Chemical Division of ASQC, conducted a survey to determine the nature and extent to which statistical control techniques were used in the food industry. Out of about 200 firms who returned questionnaires, usable detailed information was submitted by 176. Results of this survey were reported by Hosking (10) at the 1955 National Convention of the ASQC. Figure 1 graphically shows the relative frequency with which SQC was used in various food industries. Hosking in his summary stated that very little statistical quality control was employed by food manufacturers prior to 1947 but that after this date increased use was made of SQC. He reported that of the companies then using SQC about 1/2 of them had begun after 1950.

The principal application of SQC techniques in food processing has been in package fill and weight control. Hosking (1) reported that SQC was most frequently used in the packaging, manufacturing and purchasing operations in that order, and that sampling tables and control charts were the most widely used devices.

SQC IN FOOD INDUSTRIES

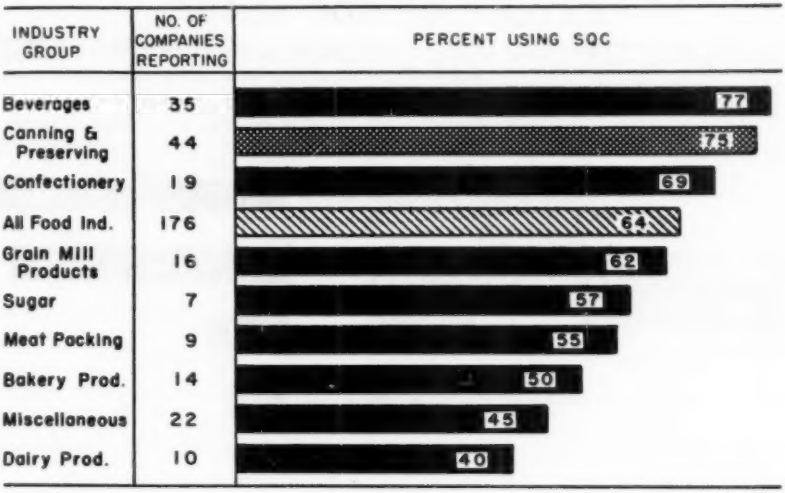


Figure 1.

In a detailed analysis of the application of SQC in the canning industry by Stier (22) it was found that in the companies where SQC was used, it was used most frequently in the packaging and manufacturing operations. The extent to which SQC was used in four areas of the canning industry operation is reflected in Figures 2 and 3.

Usually one of the first applications of SQC by a food processor is in the control of container fill (either in volume or weight). This is an application that usually pays dividends and the economic gains can readily be assessed by management. When variability of container fill is high, increased yields per ton of raw product amounting to as much as 15 percent or even higher have been recorded after using control charts from the filling line. In addition to the reduction in average weight which is frequently possible, the uniformity of fill is invariably markedly increased. Anderson (1) reports a typical example of improved uniformity before and after SQC. The examples shown in Figures 4 and 5 were taken from MacPhail (19) and represent a typical response from SQC in the filling operation. In Figure 4 not only was uniformity of fill improved but there was a five percent decrease in the average weight. This type of change would result in a five percent gain in the number of cans produced per given weight of raw material. The situation reflected in Figure 5 reflects little gain in average weight control but a marked improvement in the elimination of over weight containers and in the proportion below label weight.

CANNING AND PRESERVING INDUSTRY

EXTENT OF SQC USAGE
MANUFACTURING AND PACKAGING OPERATIONS

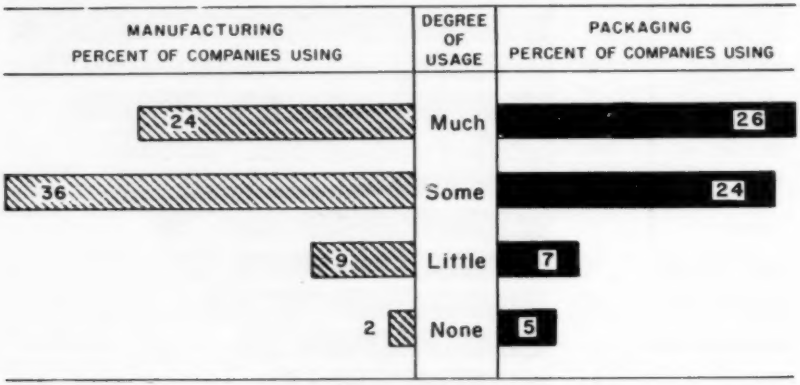


Figure 2.

CANNING AND PRESERVING INDUSTRY

EXTENT OF SQC USAGE
PURCHASING AND OTHER OPERATIONS

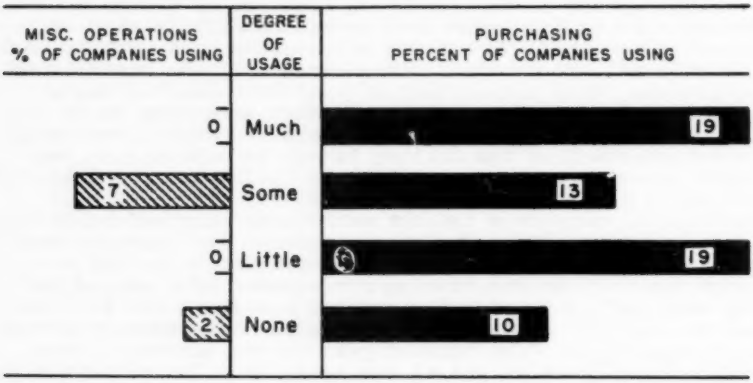


Figure 3

CANNING AND PRESERVING INDUSTRY

EXAMPLE I: IMPROVED WEIGHT CONTROL BY USING SQC

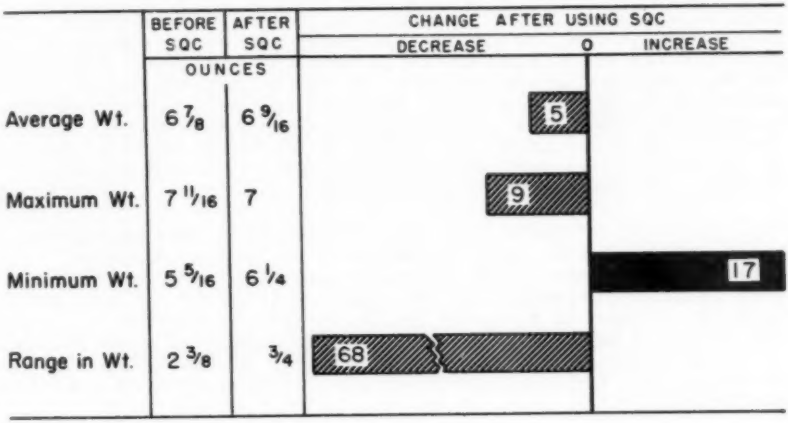


Figure 4.

CANNING AND PRESERVING INDUSTRY

EXAMPLE 2: IMPROVEMENT IN UNIFORMITY OF CONTAINER FILL BY USING SQC

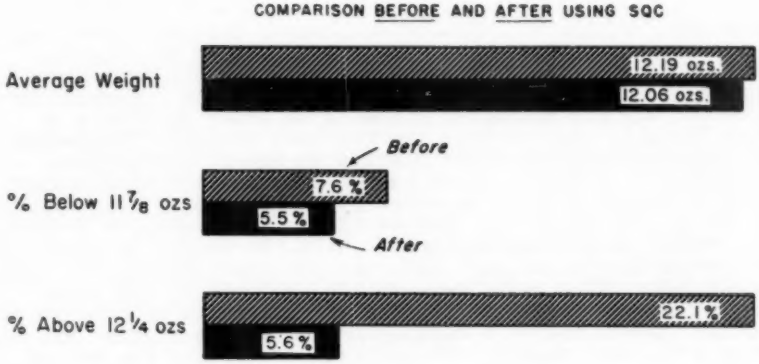


Figure 5.

Typically, control charts result in the greatest improvement in the uniformity of container fill and for gains in case yields when the filling is done by hand rather than machine. Increases of as much as 12 to 15 percent in the case yields per ton of raw product have been reported for such hand filled canned products as sauerkraut and whole green beans. However, lest the unitiated become overly enthusiastic about such potential economic gains it must be realized that in machine filled operations where the filler is in good condition, properly adjusted and a fairly up-to-date model, the uniformity of container fill will invariably be quite good. It will, for obvious reasons, be better for liquid products than for semi-solid or granular products on the same filling machine.

Way (29) presents some very useful data concerning relative precision of filling machines in filling various sizes of containers with various types of products. For a very helpful discussion of the factors involved in fill control the reader is referred to the papers by MacPhail (19), Way (28) and Edwards (7). For the techniques of controlling container fill by SQC methods the papers of Way (29), Harrison (9), Lee (16) and Gieseker (8) will provide valuable assistance.

In order to provide food canners who wished to start a container fill control program with a standard chart that they could use for $\bar{X}$ - R charts, the National Canners Association printed in quantity blank chart forms. The chart in Figure 6 illustrates a typical $\bar{X}$ - R chart drawn on the NCA form which is provided as a service to canners who wish to initiate a fill control program.

In addition to the use of $\bar{X}$ - R charts in container fill control, other applications by food processors have included charting raw product quality, quality of incoming container shipments, equipment evaluation, product quality, labeling, and processing efficiency. Control charts on raw product quality provide the food processor with a basis for determining the appropriate production usage of the incoming raw material and the basis for blending in order to attain a given quality attribute in the finished product. Shipments of incoming containers are sampled in order to determine adherence to specifications and eliminate a possible source of deterioration in product quality. For a good discussion of the practices and techniques of container quality control, the reader should become familar with the papers by Lang (15) and Weimerskirch and Way (32). In the paper by Way (29) and by Gieseker (8) a brief description is given of the procedures involved in evaluating the effectiveness of fillers at various speeds and with various products. In a paper by Stier (24) some suggestions are made concerning the use of SQC techniques in evaluating the effectiveness of sanitation operations in the food processing plant. Other applications that have been made in the food processing industry include control of certain chemical and physical attributes.

STATISTICAL QUALITY CONTROL CHART
Pill Weight - Ounces
Plant 7
Container size - No. 2½ Can
Date August 18-28
Product Canned Peaches
Shift
Code
Averages
Ranges
UCL
LCL
Time

Figure 6.

Food canners who are using statistical quality control techniques were asked in a survey reported by Stier (22 and 23) to state the benefits they obtained from using SQC. The response from 44 different canners of various sizes is tabulated and graphically shown in Figure 7. Most frequently mentioned benefits were higher quality products, better weight and fill control, aided in meeting customer specifications and fewer failures or rejections.

Those food processors who reported very little usage of SQC were asked for the reasons for such limited use. The summary of the response is graphically shown in Figure 8. The high percentage of companies reporting they did not feel SQC was applicable to their operation reflects a general lack of imagination in modification of SQC procedures to food processing or an incomplete understanding of the technique itself. The mention of personnel shortage indicates an awareness of the lack of the necessary training by persons in the plant.

CANNING AND PRESERVING INDUSTRY

Figure 7.

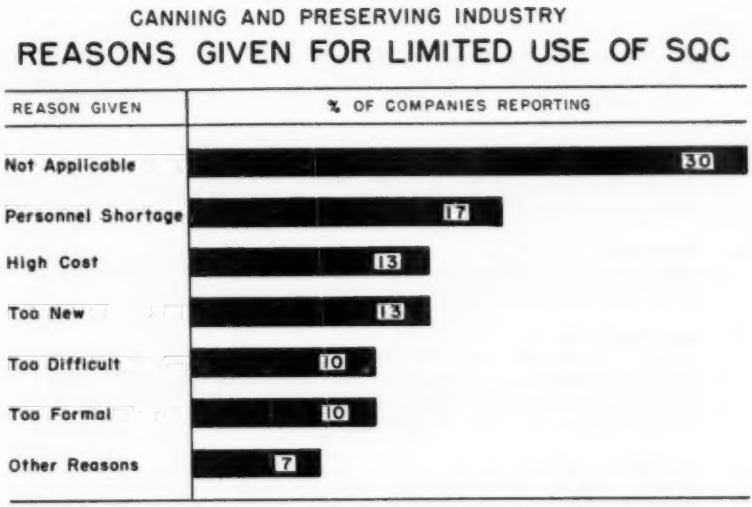


Figure 8.

There are a few instances of companies in the food processing field familar to the author where SQC is used in a highly integrated operation. Scientific sampling is followed in the selection of raw product for certain grades or packs of the finished product; quality control charts are maintained of various physical and chemical qualities important to product quality and package or container fill is controlled by periodic sampling and $\bar{X}$ - R charts. Although this type of application is extremely effective and apparently returns more than it costs the frequency of such application is very low. The development of efficient physical measures of measuring quality is providing more and more opportunities for the use of the correlative measurements that speed up the application of SQC in food processing.

Other areas that would seem to be worth serious consideration and trial are the use of SQC techniques in controlling quality of purchase supplies, accident control, cost control, production worker efficiency and the coordination of market trends and marketing data with production plans and operations.

The list of references which follows is not intended to be comprehensive but rather to include those papers which might provide those persons in the field of food processing who are interested in applying SQC with various points of view and a description of techniques used to be helpful in initiating a program.

References

1. Anderson, L. P. & R. D. Kornblum
1957. Semantics, Scales, and Savings Through S.Q.C. 1957 ASQC National Convention Transactions:379-384

2. Bayton, James A.
1955. Rating Scales and Psychological Factors In Taste Preference Research, 1955 ASQC National Convention Transactions:275-282

3. Bradley, Ralph A.
1955. Statistical Designs for Taste Test Panels, 1955 ASQC National Convention Transactions:621-626

4. Burrows, G. L.
1956. Comments on the Validity of Methods Used in Taste Test Research, unpublished manuscript of paper presented at the 1956 Middle Atlantic ASQC Conference in Washington, D. C., February 10, 1956

5. Carroll, Mavis B.
1956. An Application of Paired Comparisons in Evaluating Sensory Responses, unpublished manuscript of paper presented at the 1956 Middle Atlantic ASQC Conference in Washington, D. C.

6. Dykstra, Otto, Jr.
1957. Applications of Scheefe 'S Method of Paired Comparisons in Food Quality Evaluation, 1957 ASQC National Convention Transactions:167-175

7. Edwards, H. Ivor
1954. Filling Semi-Solid Products, National Canners Association 1947 Convention Proceedings:76-77

8. Gieseker, Leonard
1955. A M_0dified Lot Plot Procedure for Controlling Container Fill. 1955 ASQC National Convention Transactions:265-273

9. Harrison, Stephen
1952. The Use of Statistics in Food Processing, 1952 ASQC Quality Control Conference Papers:121-128

10. Hosking, Floyd J.
1955. Quality Control Techniques Used In The Food Industry, 1955 ASQC National Convention Transactions: 293-297

11. Kramer, Amihud
1952. Quality Control Simplified for Practical Product Improvement, Food Engineering:100-108

References (continued)

12. Kramer, Amihud
1956. Selecting An Efficient Sampling Procedure For Food Products, 1956 ASQC National Convention Transactions:465-470

13. Kramer, Amihud
1956. The Use of Sensory Test Panels in Process Quality Control, 1956 ASQC National Convention Transactions:119

14. Kenny, Frank
1950. New Concepts of Statistical Quality Control For Food Processors' Rutgers University Quality Control Conference, September 16, 1950:75-77

15. Lang, K. F.
1957. Problems and Pitfalls in Acceptance Sampling Of Glass Containers, 1957 ASQC National Convention Transactions:119-130

16. Lee, Edward P.
1956. A Statistical Quality Control Approach to Weights and Measures Problems, 1956 ASQC National Convention Transactions:543-551

17. Lee, Edward P.
1956. Statistics in Quantity Control of Packages, Proceedings 41st National Conference on Weights and Measures

18. Mahoney, C. H., H. L. Stier and E. A. Crosby
1957. Evaluating Flavor Differences in Canned Foods, Food Technology, Vol. XI, No. 9:29-36

19. MacPhail, Ian
1954. Filling Liquid Products, National Canners Association 47th Convention Proceedings:70-71

20. Peryam, David R.
1957. Factors Affecting The Accuracy and Reliability of Sensory Tests, 1957 ASQC National Convention Transactions:675-685

21. Stier, Howard L.
1952. Some Simplified Statistical Procedures Useful in Evaluating Flavor Differences, unpublished manuscript of paper given at the annual meeting, American Association of Economic Entomologists, December 16, 1952

22. Stier, Howard L.
1955. Statistical Quality Control in the Canning and Preserving Industry. National Canners Association, October, 1955, 19 pages

References (continued

23. Stier, Howard L.
1957. Quality Control Practices and Problems in Food Processing. Conference Papers--Quality Control and the Consumer Conference, September 5-6, 1957, Rutgers University:21-31

24. Stier, Howard L.
1958. Methods of Statistical Control for Sanitation Operations. Modern Sanitation and Building Maintenance, Vol. 10--No. 6:16-18

25. Stier, Howard L.
1958. Statistical Quality Control--Milestone or Millstone? Report of 5th Annual Research Symposium, National Association of Margarine Manufacturers, Chicago, Illinois, October, 1958

26. Stier, Howard L.
1959. Effective Use of Cost Data for Plant Operations. Proceedings, 9th Annual Food Processors Workshop, University of Maryland, College Park, Maryland

27. Tenney, Robert I. and Marjorie L. Sutherland
1956. Statistics Sharpen the Brewers Quality Control Tools. 1956 ASQC National Convention Transactions: 389-396

28. Way, C. B.
1954. Fill Container of Granular Products. Proceedings National Canners Association 47th Annual Convention: 73-76

29. Way, C. B.
1955. Fill Control in the Canning Industry, 1955 ASQC National Convention Transactions:505-512

30. Way, C. B.
1959. How Canners Can Use Statistical Quality Control. Spring Meeting, Wisconsin Canners Association--Processors Conference, March, 1959

31. Weiss, William
1956. Statistical Considerations in Food Standards. Proceedings, NCA Statistical Quality Control Conference, April, 1956

32. Weimerskirch, R. I. and C. B. Way
1956. A Mutual Approach To Quality Control by Can Manufacturer and Food Processor. 1956 ASQC National Convention Transactions:313-321

BULK SAMPLING

W. M. Bertholf
The Colorado Fuel & Iron Corp.,
Pueblo, Colorado.

INTRODUCTION.

"Few subjects have been so frequently discussed as ore sampling. The discussion, whilst apparently endless, is very necessary. It seems impossible to fix in the mind of the mining public the very simple laws that govern the operation of sampling. No operation is easier to perform with certainty and accuracy than that of sampling."

A. W. Warwick (1902) (1)

Mr. Warwick was speaking of the determination of the value of individual lots of ore to the mutual satisfaction of seller and buyer at a time when it was estimated that the mineral production of Colorado alone was about $50 million per year. Small wonder that there was discussion.

It is, however, to be doubted that all bulk sampling is inherently as easy as Mr. Warwick made it appear. A careful study of the 40-page booklet indicates that there are several major rules or laws governing the collection of a satisfactory ore sample.

1. In order to properly obtain a sample of ore, it is necessary to take the same quantity of ore frequently, or in as many places as possible, and at regular intervals.

2. The size of the sample (to be taken at any stage) varies as the cube of the largest piece.

3. A good sampler need not mix the ore.

4. The salient points concerning streams of ore are:
   a) the ore stream changes gradually from rich to poor and vice versa in the direction of flow
   b) the distribution of values across the stream is extremely irregular, and the different parts of the stream are vastly different, hence

5. A sampler which takes a transverse slice of the stream must take a true cross-section of the stream, and

6. A sampler which takes a (fixed) longitudital slice of the stream cannot possibly take a correct sample.

7. The minimum opening of the sampler must be great enough to allow the largest pieces of ore in the falling stream to pass through the opening with absolute freedom.

8. During the sampling operation the stream should be delivered steadily to the sampler and in as solid a condition as possible.

It is concluded that if the machine is simple, readily cleaned, will stand much wear and tear, does not require to be stopped at frequent intervals for adjustment or repair and conforms to the basic principles given above it will certainly give a most accurate sample in the quickest possible time, and at a very small expense.

In essense, the story hasn't changed a bit since 1902. Later discussions of the problem are sometimes more elaborate but seldom are they fundamentally better.

It is unfortunate that Mr. Warwick included no direct evidence to prove that a sampling machine actually takes a good sample—all his evidence is highly circumstantial, involving comparisons of actual mill recoveries with metallic content of the feed estimated on the basis of in-the-mill sampling, not samples of incoming ore. However, the testing of a sampler is touched upon, and I quote,

> "In order for a test of a sampling machine to be complete, it should include a set of screen analyses besides assays. It goes without saying that the sample of the rejected ore should be made by the method of taking alternate shovelfuls. We lay great stress upon this method of testing since it demonstrates, in conjunction with the assays, not only whether a machine is capable of taking a good sample, but also another very important point, viz, whether the cuts are being taken often enough. For example, if we find that the screen analyses of the sample and of the rejected ore are the same, or practically so, but that the assays do not agree as closely as they should, then we see no other possible explanation than that the cut has not been taken frequently enough and hence a long portion of rich or poor ore, as the case might be, has got through the machine without being sampled. A machine that is perfectly good in principle might be condemned if the method of checking by assays only is adopted. The fault might be that the cuts are taken too far apart."

The "utter simplicity" of the machine sampling of ore of fifty years ago was partly real and partly illusion. Considering the reality of the problem there is no doubt but that some method which was virtually foolproof was essential. The vendor could not be trusted to bring in a sample which could be used as a basis for payment—and the buyer's methods of determining value were likewise under a cloud of suspicion—the only way out of the dilemma was to develop a system which would be mutually agreeable. With much of the ore running $200 to $300 per ton this system need not actually be simple nor cheap.

The time-honored system was to "cone and quarter" the lot. This involved shoveling the lot into a cone, flattening, taking alternate quarters as the sample, crushing to a smaller size (say to one-half) and repeating the process ad nauseum. This was slow and expensive. A considerable reduction in time and expense was achieved by the "alternate shovel" method, which was capable of considerable abbreviation by taking every n-th shovel, n sometimes being as high as ten.

Today we find millions of tons of material with a value of $5 to $10 per ton being bought on an analysis basis—the total value being much greater than that of the gold and silver ores of fifty years ago. Can we or should we expect to get results as fantastically close as is claimed for the ore sampling operations? Warwick cites "check" samples agreeing within 40¢ per ton on $200 ore. This would mean 1¢ per ton on $5 ore.

Present indications are that we are unlikely to obtain a precision of sampling and analysis of this magnitude on single lots without a drastic overhauling of the entire sampling process. None of the data with which the writer is familiar indicate that the customary procedures are even close to the mark, if, indeed, they are aiming at it.

Pasek and Lundberg (2) have indicated that it is possible to estimate the Btu content of dry coal with a standard error of 25 to 30 Btu in the 11,000 to 12,000 Btu range. Other data derived from Round Robin tests (3) indicate that the accuracy of analysis on the "as received" basis (with about 1.5% moisture in the 60-mesh pulps) is in this range; but on 8-mesh samples (with about 6.5% moisture) the standard error is 50 Btu or more. Further, while the relationship "rediscovered" by Pasek and Lundberg appears to hold within laboratories it will not hold on inter-laboratory comparisons. There are reel differences between laboratories.

Much the same could be said for moisture, ash, sulfur, volatile matter, etc. in coal. These determinations are purely empirical, i.e. they are not based on chemical reactions as such, carried out under conditions which insure unambiguous results.

In coal sampling there is a concurrent search for ways of improving both sampling per se and the methods of analysis. Sub-committees of ASTM Committee D-5 are actively engaged in this work.

TEST DATA ON SAMPLERS.

Our first example was reported by Bertholf and Webb (4). Those interested in complete details will have to refer to the original paper, but we would like to emphasize a few items in the light of Warwick's analysis of the situation.

At Cabin Creek the problem was to obtain an adequate sample of a flow of coal on the order of 750 tons per hour. It is immediately apparent that such a flow rate puts the problem in a realm far removed from that of the ore sampling of 1900. Nevertheless it is necessary to get a good sample.

Applying Warwick's first test, in a slightly modified form, comparisons were made of the size of material on the belt and that taken by the sampler. Are the screen tests "practically identical"? See "A" in Fig. 1. The analysis of variance of the data indicates that there is a significant difference between the two sets, with the sample slightly smaller than the feed. "F" is slightly above the 5% point.

Our opinion is that this is to be expected with an "opening ratio" approximately 3:1. It is not likely that the full quota of large pieces will get through an opening three times as large as the "largest pieces"

but if the values are not concentrated in the largest pieces the effect on the sample will be negligible since there are only a limited number of largest pieces. In this case we had approximately 2% of the coal in the plus 1½-in. size. Apparently only 95% of the expected number of such pieces reported in the sample. The computed difference in ash analysis due to this effect is about 0.01%. It is virtually impossible to isolate an error of this magnitude.

Rather than compare the results of daily composites the ash comparisons were made on a chronological basis. Fig. 2 is taken from the original paper and presents the running record of three sets of samples. The "hand" increment average is the weighted average of the 15-, 40- and 100-lb. increments taken at each "hourly" stopping of the belt. The "parallel" machine increments were taken at approximately the same time as the hand increments and the "alternate" machine increments were taken about midway between the hand sets.

This is an excellent set of data for someone interested in the analysis of non-stationary time series and confirms Warwick's observation that the stream gradually changes from rich to poor and vice versa in the direction of flow.

Since the sampler made a cut completely across the stream no data were accumulated on the transverse cross-section.

We believe that the Cabin Creek results demonstrated that a sampler can take a sample which will adequately represent the stream in almost all respects. If one insists upon "identical correspondence" between sample and main stream he must be prepared to pay the price, but there seems to be no theoretical barrier to reducing the real difference to a value which is of no practical importance if the design of the machine is sound and the material will flow through it.

At the same meeting, Blatter (5) reported on a slotted revolving cylinder. His test was not as elaborate as the Cabin Creek test but was extensive enough to demonstrate that there is a difference between samplers. In this case we had 11 sets of samples consisting of five samples taken from a stopped belt and five cuts taken with the "spoon" in rapid succession. In both cases the five samples were spread across the belt but the spoon samples were also spread along the belt--adding some to the difficulty of interpretation of the results.

Fortunately certain critical items are so clearly demonstrated that it is unnecessary to split hairs about the others.

The opening through which the sample is collected in this type of sampler is normally down and the flow of coal goes around the cylinder. At regular intervals a timer revolves the cylinder and for a portion of the complete revolution the slot collects a sample which is diverted out of the main flow. The cylinder is continually traversing the belt and over a period of time will collect samples from every part of the stream.

Knowing the typical cross-section of the belt load, the rate of revolution of the cylinder and the width of the opening, it is possible to compute the amount of material that should be collected at any point in the traverse. The following tabulation indicates that only on the outside

positions (1 and 5) did the sampler take the theoretical poundage.

Table 1.

Position	Pounds Coal per Cut	
	Theoretical	Actual
1	22.2	22.4
2	36.8	24.8
3	48.5	27.4
4	37.4	27.6
5	20.4	25.5

Obviously this sampler overloads the sample with material from the sides of the belt and is definitely deficient in the heavily loaded center.

It is no wonder that weighted averages of the automatic samples did not agree with weighted averages of the hand samples, for there were real differences in the material from different portions of the cross-section of the belt, as evidenced by the analysis of variance of the ash data for both hand and machine samples.

Blatter's original presentation of the comparative size of material taken from the different positions by hand and machine is a bit more vivid than our comparison of the over-all results in Fig. 1, "B". The revolving cylinder definitely was biased. In our opinion any attempt to reduce the size bias by widening the slot would simply increase the bias indicated in Table 1, which we believe is due to choke feeding of the cylinder and would be just that much more noticeable since the tube thru which the sample flows would remain the same but would be expected to carry more.

The two examples indicate, to us at least, that machine sampling can be good or bad, depending on the design characteristics of the machine and what it is used for. When it is good it yields information which would be prohibitively costly otherwise. When it is bad one might be just as well off without it. On the other hand, most of us use figures which are consistently biased to a certain extent, and an automatic sampler that is not entirely perfect would undoubtedly furnish information that could be used for action--and that's what it is all about anyway.

The following data are taken from my files and are used simply to illustrate a rather free and easy interpretation of one criterion for the usefulness of samples, viz, if you get the information you need to act intelligently the samples are useful.

THE SAMPLING OF RAW COALS.

We receive five different coals in railroad cars. In order to keep the producers "on their toes" it is essential that we keep a running record of the quality of each coal. It must be good enough to detect real changes in quality and consistent in the sense that any time we complain we will not be sticking our necks out over nothing at all.

Not long ago we changed our raw coal sampling from the car tops to the individual mixing belts. Now that we are able to put each coal

FIGURE 1 - COMPARISON OF SCREEN TESTS - SAMPLE AND STREAM

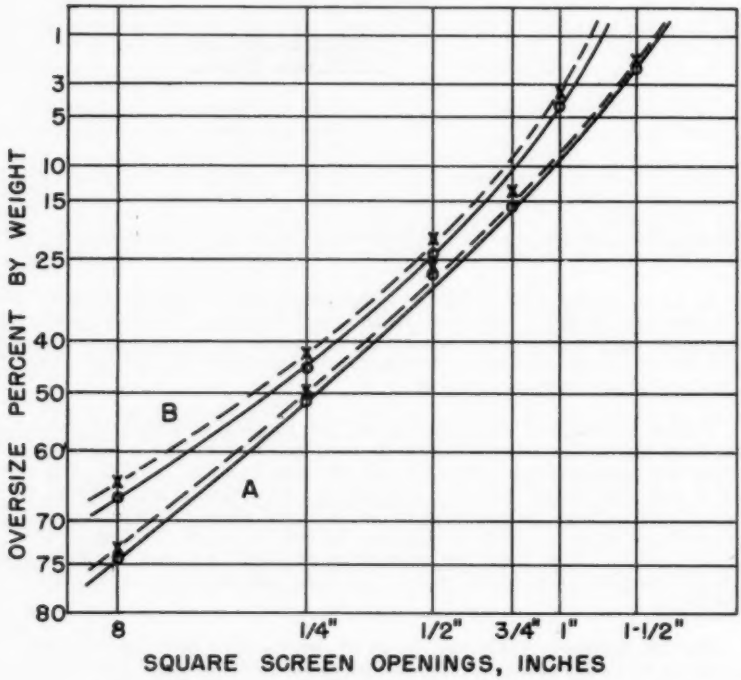
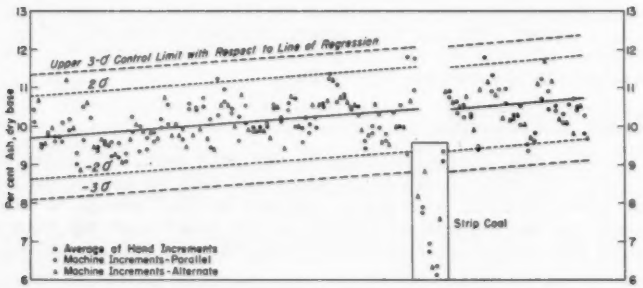


FIGURE 2 - CHRONOLOGICAL PLOT - CABIN CREEK SAMPLES



through its own bin, this appears to be a logical move. At least one of our coals should get a "break" from this change since it is layer-loaded with the bad material on top. Another might be expected to agree more closely with samples taken during loading since there is a very strong probability that segregation develops during an 800 mile rail haul. The effect of the change on the other three coals was not expected to be more than barely noticeable, since all are relatively fine and loaded rapidly.

Tables II and III show averages and standard deviations for a selected group of determinations normally made on weekly composite samples of the individual coals. There are, of course, many other determinations that could have been tabulated. In all cases we assembled the weekly data for a coal in two columns "cars" and "belt". The car samples are from November and December, 1958, the belt samples from January and February, 1959. A standard analysis of variance procedure was used for the one-way classification.

It will be noted that the volatile matter, ash and float-sink of FCL and Arky changed significantly as to average but there was no noticeable change in standard deviation. This had been anticipated, and is tentatively assumed to be due to the changed sampling procedure.

Since FSS has been going through some gyrations of its own in the recent past we are more inclined to believe that the observed changes in volatile and ash are due to changed conditions in the mine than due to the change in sampling procedure.

As expected, there is no noticeable difference in the Allens.

This experience tends to confirm the opinion, shared by at least a few of my friends, that car-top sampling is a "sometimes" thing. Sometimes it's fine, other times it's a waste of money. We doubt that it takes a genius to tell whether car-top sampling is going to work out. One good look at the loading system should be enough.

While the standard errors of the weekly samples for individual coals are rather high the composite monthly average, which is used to determine washery efficiency, etc. is reasonably reliable, even if it is no better than is indicated in the column at the extreme right of Table III.

In all probability the reliability is even greater than indicated. In Fig. 3 we show the running record of the monthly average ash from Jan., 1955 to date. Through this array of spots we have run relatively smooth curves representing the "trend". This situation is very similar to that found at Cabin Creek, Fig. 2.

During 1955-56 we had a relatively smooth sine-wave fluctuation in ash and the monthly averages stayed within 0.5% of this base curve. The r-m-s deviation from the curve was 0.25%. Starting in 1957 the pattern broke, and one is quite certain that number of the recent deviations from trend are not due entirely to sampling errors. The last nine points on the chart show a very considerable decrease in ash following an abrupt increase and are in line with certain changes in mining practice. One must admit that if this pattern is not due to something real it is a very unusual sequence of errors.

TABLE II. Averages of Weekly Samples

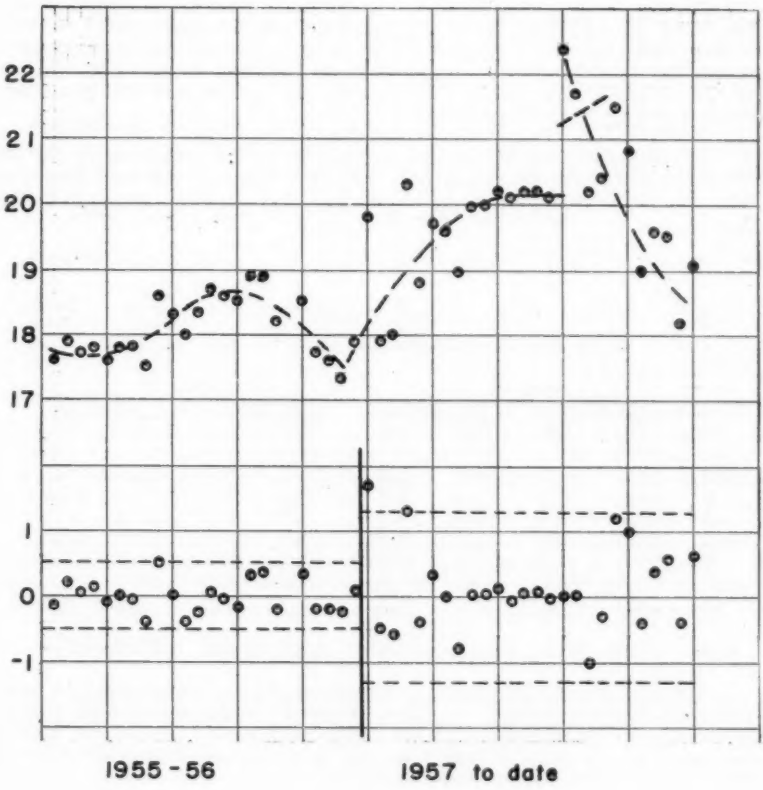
Constituent	From	FSS	FCL	E.All	W.All	Arky
Vol. Matter	Cars	31.22	26.44	34.77	36.03	16.61
	Belt	32.34*	28.44*	34.85	36.26	17.20*
Moisture	Cars	3.30	4.15	4.19	4.02	4.26
	Belt	3.79	3.35	5.37*	4.43	4.64
Ash	Cars	18.62	33.80	18.72	19.00	9.48
	Belt	16.69*	27.88*	18.98	19.72	8.22*
Fl. @ 1.38	Cars	79.40	47.16	79.85	80.25	85.87
	Belt	79.48	56.16*	79.88	79.18	89.21*
Heavy Bone	Cars	1.04	3.24	0.88	0.81	0.45
	Belt	1.19	3.66	0.59	0.79	0.35
Rock	Cars	14.50	35.53	15.35	16.60	6.13
	Belt	12.70	26.98*	15.90	16.28	3.71*

* Statistically significant change in average quality.

TABLE III. Standard Deviations of Weekly Samples

Constituent	From	FSS	FCL	E.All	W.All	Arky	Monthly Composite
Vol. Matter	Cars	0.52	0.47	0.80	1.21	0.41	0.21
	Belt	0.33	0.62	1.18	0.71	0.17*	0.21
Moisture	Cars	1.04	1.67	1.49	0.82	1.20	0.30
	Belt	1.05	1.07	1.10	1.18	1.17	0.26
Ash	Cars	2.10	1.67	1.48	2.55	1.20	0.47
	Belt	0.25*	1.64	1.92	1.50	0.62	0.36
Fl. @ 1.38	Cars	2.94	3.94	2.80	2.95	3.11	0.71
	Belt	0.79*	3.32	3.31	3.31	2.14	0.69
Heavy Bone	Cars	1.02	1.54	1.14	0.62	0.51	0.23
	Belt	0.81	1.53	0.69	0.90	0.62	0.20
Rock	Cars	3.02	2.70	2.36	2.96	1.74	0.63
	Belt	1.24*	2.73	3.15	2.81	1.25	0.63

FIGURE 3 - MONTHLY AVERAGE ASH - RAW COAL



THE SAMPLING OF WASHED COAL.

The sampling of washed coal is much easier than the sampling of raw coal of comparable size, and in an operation such as ours where the washed coal is crushed before it is sampled there would appear to be no valid reason for not getting useful samples by almost any method one might care to use. Such is not necessarily the case.

When washed coal was sampled from the belt leaving the washery we obtained erratic results all too frequently. This belt is actually layer loaded with up to five kinds of stuff, including a layer of pitch on the bottom. With a sharp increase in washery throughput after installing new equipment it actually was dangerous to stick a shovel into the stream to get a sample. The sampling point was moved to two belts involved in making the transfer from a storage bin to the oven bins. This took care of the safety hazard but, unfortunately, did not improve the character of the laboratory reports noticeably.

Several years we made an investigation of the situation at "I" and "J" belts and found by taking full cross sections of the belt load that this was probably a poor place to sample. The data are shown in Table IV.

TABLE IV. H ourly Samples from "I" and "J" Belts.

	Jan. 24				Jan. 30			
	"I" Belt		"J" Belt		"I" Belt		"J" Belt	
	H_2O	Ash	H_2O	Ash	H_2O	Ash	H_2O	Ash
1	12.6	9.4	15.0	9.8	11.1	8.6	14.4	9.3
2	11.0	8.3	13.1	8.7	7.3	9.3	19.1	10.7
3	11.0	9.7	14.0	9.5	7.8	9.9	11.9	10.5
4	18.9	9.6	11.1	9.5	7.3	9.6	12.2	10.4
5	11.3	8.9	11.6	8.9	12.3	9.0	11.6	8.9
6	6.8	9.1	16.3	10.0	13.1	9.3	12.9	9.1
7	13.3	9.1	13.0	9.5	12.3	10.0	8.4	9.6
Avg.	12.1	9.1	13.4	9.4	10.2	9.4	12.9	9.8

One need not be an expert statistician to see that this is a very unsatisfactory situation. At first glance it would appear that there is no hope of straightening it out. The only reason we did not give up was that we knew that it wasn't as bad as it looked--there was some way of getting better samples than these.

We transferred our attention to the remaining point where the truth might be found, if we were ever to find it, with the following results. Small samples were taken from the top of the larry car, one canfull from each of the four hoppers. These were checked for size, moisture and ash. In Table V we show only the moisture and ash data--screen size data is more voluminous than we have space for, but is about the same calibre as those data we include here.

TABLE V. Moisture at Larry Car Hoppers, North Coal Bin.

January 31					February 7				
Hopper					Hopper				
1	2	3	4	All	1	2	3	4	All
14.0	13.2	15.9	14.3	14.3	13.9	15.8	13.0	13.5	14.1
11.9	13.5	12.3	13.2	12.8	12.3	13.8	13.5	13.7	13.3
13.3	12.9	13.0	13.7	13.2	13.8	14.5	13.7	14.2	14.1
12.9	12.2	10.5	12.9	12.1	15.0	13.7	17.4	13.6	14.9
13.6	11.9	11.9	11.8	12.3	14.0	(14.6)	15.5	14.3	14.6
13.8	13.0	13.9	13.4	13.5	14.8	14.0	15.9	13.7	14.6
					16.6	14.6	15.6	14.5	15.3
13.3	12.8	12.9	13.2	13.0	14.3	14.4	14.9	13.9	14.4

Analysis of variance on the two-way classification was expected to show considerable difference from day to day, possibly from hour to hour and, we hoped, none from hopper to hopper. It turned out as in Table VI.

TABLE VI. Analysis of Variance, Moisture Data.

January 31.				February 7.			
Source	S.S.	d.f.	M.S.	Source	S.S.	d.f.	M.S.
Rows	13.88	5	2.78*	Rows	10.34	6	1.72
Cols.	0.94	3	0.31	Cols	3.64	3	1.21
Rem.	11.60	15	0.77	Rem.	19.16	18	1.06

It appeared that we would be in good shape if we took a composite sample, one increment from each hopper of the larry car, once an hour. The "error" of the average of 16 such samples, two shifts per day, would not be expected to exceed 0.5% very often. This would be a very definite improvement over the old situation.

Corresponding data for Ash are given in Tables VII and VIII. Note the decrease in error mean square for ash as compared to moisture in washed coal. In the raw coals moisture was the more uniform of the two determinations

TABLE VII. Ash at Larry Car Hoppers, North Coal Bin

January 31					February 7				
Hopper					Hopper				
1	2	3	4	All	1	2	3	4	All
9.3	9.4	9.6	9.4	9.4	9.9	9.9	9.8	9.9	9.9
9.0	9.3	9.2	9.3	9.2	9.8	9.5	9.8	9.8	9.7
9.2	9.3	9.7	9.6	9.4	9.2	9.8	9.7	9.5	9.6
9.6	9.3	9.3	9.6	9.5	9.3	9.3	9.7	9.1	9.4
9.1	9.1	9.5	9.1	9.2	9.6	9.4	9.6	9.6	9.5
9.4	9.4	9.7	9.6	9.5	10.6	10.1	10.2	10.1	10.2
					10.1	10.0	10.3	10.3	10.2
9.2	9.3	9.5	9.4	9.4	9.8	9.7	9.9	9.8	9.8

TABLE VIII. Analysis of Variance, Ash data

January 31				February 7			
Source	S.S.	d.f.	M.S.	Source	S.S.	d.f.	M.S.
Rows	0.393	5	.0786*	Rows	2.72	6	0.45**
Cols.	.218	3	.0706	Cols.	.09	3	.03
Rem.	.334	15	.0223	Rem.	.65	18	.036

Following the same line of reasoning as was used on the moisture, the "error" of a daily average ash should be approximately 0.03%, but this assumes that the composites can be worked down for the laboratory without error and that the laboratory work will be perfect—neither of which is a good working basis. There will always be some measurement error, and in this type of material it may well be several times the true sampling error.

THE ERRORS OF REDUCTION AND ANALYSIS.

It is usually assumed that the errors of reduction and analysis are small compared to the sampling error. This may be true in some cases, but with a sampling error as small as we found for the ash in washed coal it would be well to check this point before it is too late.

The alternate methods of preparing an analysis sample from the hourly larry sample are: 1) split out a small portion and grind to (say) 50 mesh, 2) grind the sample to 6-mesh, then split and pulverize or 3) pulverize the entire sample and take out the 20-30 g. for the laboratory. In our book procedure (3) would be used only as a last resort, so we checked procedures (1) and (2).

A batch of washed coal, approximately the same weight as a normal hourly sample was riffled into 16 portions, each of which was pulverized and reduced to 20-30 grams for analysis. This set is Series I. Another batch was ground to pass 6-mesh before the riffling—Series II. Table IX gives the results.

TABLE IX. A Comparison of Two Methods of Sample Reduction.

Split	Series I	Series II	Split	Series I	Series II
1	9.2	9.2	9	8.7	9.1
2	9.2	9.3	10	9.2	9.1
3	9.2	9.1	11	9.4	9.1
4	9.0	9.2	12	9.1	9.2
5	9.1	9.1	13	8.7	9.0
6	8.2	9.1	14	8.4	9.0
7	8.6	9.1	15	8.7	9.1
8	8.8	8.9	16	8.8	9.1
			Avg.	8.9	9.1
			s^2	0.11	0.0087
			s	0.33	0.093

It is obvious that to use procedure (1) amounts to throwing away all the gains we made by finding the right place to sample washed coal. Procedure (2) is good enough for our needs, and we have never gotten around to finding out what happens with procedure (3).

Time does not permit going into great detail on this phase of the sampling process, but it seems clear to us that this is a very fertile field for miniature automatic samplers. After the sample has been air-dried one will have no difficulty in maintaining flow through them--and, if properly designed and operated, they are uncannily accurate. We have a pair of Rube Goldberg's which our samplers consider to be worth their weight in gold.

CONCLUSION.

We are happy to report that the number of "false alarms" from the laboratory reports has diminished considerably since our investigation of the sampling problem.

One rather disconcerting result of the increased efficiency of the sampling program has been the tendency of the oven operators to complain bitterly about variations in moisture content of the washed coal that a few years ago were not even noticed. In 1954 the standard deviation of a daily moisture in washed coal, based on the day-to-day changes, was 1.05%. In 1959 it is 0.33% -- and they squawk!

We believe that we have shown that at least some bulk materials can be sampled satisfactorily if proper precautions are taken to insure that the sampling point is such that an unbiased sample can be taken, the samples are taken regularly and, above all, treated with the respect to which they are entitled.

The bulk sampling problem cannot be solved without coordinated effort along three lines: 1) careful analysis of the physical set-up, 2) proper treatment of the samples after they are taken and 3) a follow-up to see whether there is any real basis for the troubles indicated by the samples. It will be necessary to keep going around the circle from now on--there is no getting off place.

REFERENCES.

1. A..W. Warwick, "Notes on Sampling", The Industrial Printing and Publishing Co., Denver, Colo.
2. R. L. Pasek and R. M. Lundberg, "Determination of Heat Content of coal by Regression Analysis," Paper 56-FU-4, ASME,
3. Private communication from T. A. Miskimen and P. Fisher, American Electric Power Service Corporation.
4. W. M. Bertholf and W. L. Webb, "Tests of the Geary-Jennings Sampler at Cabin Creek," ASTM STP No. 162,
5. A. O. Blatter, "A Test on a Slotted Revolving Cylinder Coal Sampler," ASTM STP No. 162.

SUGGESTIONS FOR THE CONTRIBUTIONS THAT THE DATAPLOTTER MAY MAKE TO THE IMPLEMENTING OF MODERN DATA PROCESSING TECHNIQUES TO STATISTICAL QUALITY CONTROL

James H. Smiley
Standard Instruments
Electronic Associates, Inc.

With the advent of the digital computer, extremely useful data on every conceivable facet of modern business can be easily and quickly generated. The Dataplotter is a data reduction system and goes hand-in-hand with the computer to reduce this data to a form in which it is easily assimulated. It takes the information from punched cards, punched tape or magnetic tape and reduces it to visual form.

The Dataplotter performs the same function in producing charts and graphs as the computer does for data handling. That is, it eliminates the time and expense involved in producing the desired result. Two main features of the Dataplotter are accuracy and speed of plotting. The Dataplotter has an accuracy of .05% full scale; on the 30" x 30" plotting surface this would mean the point is within .015" of the value described by the recorded data. Moreover, the Dataplotter, being a machine, does not tire as a human does, its accuracy remains constant.

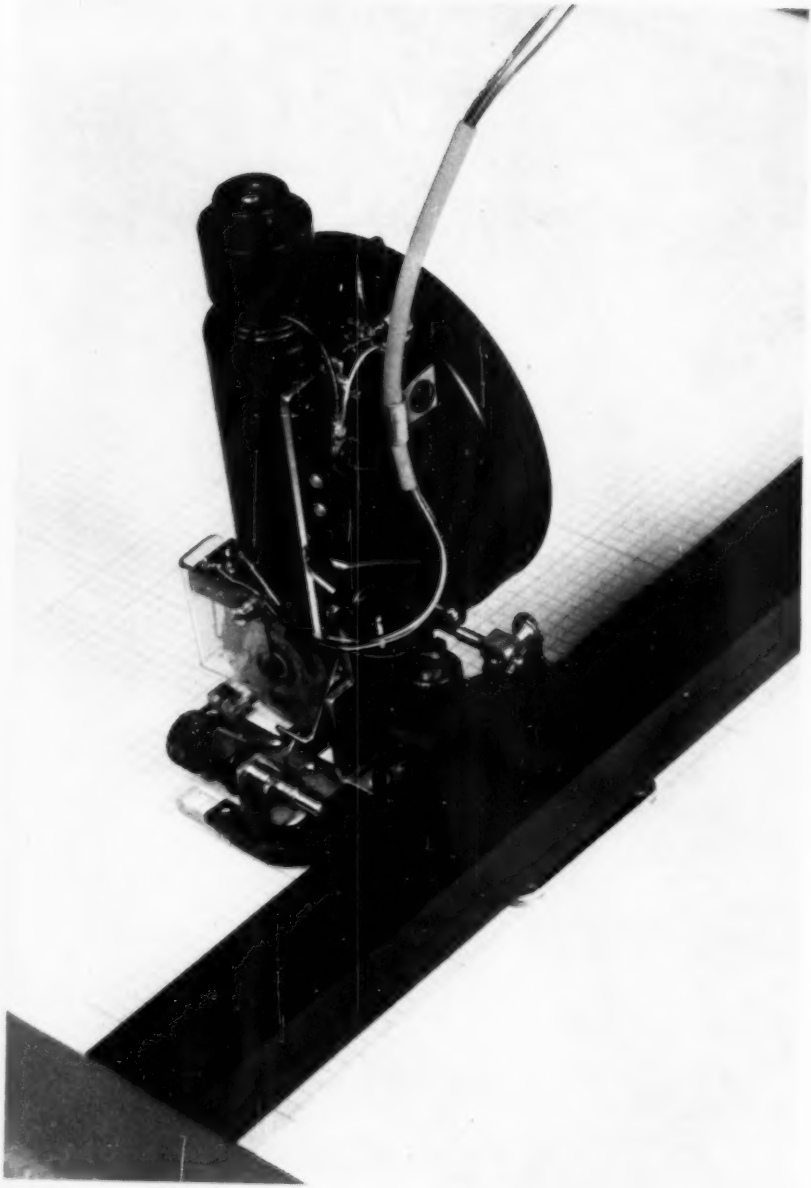
Preparation of the data for the plotter is a computer function. The information can be stored in whatever form of output the computer has (cards, punched tape, magnetic tape) for plotting independently of the computer.

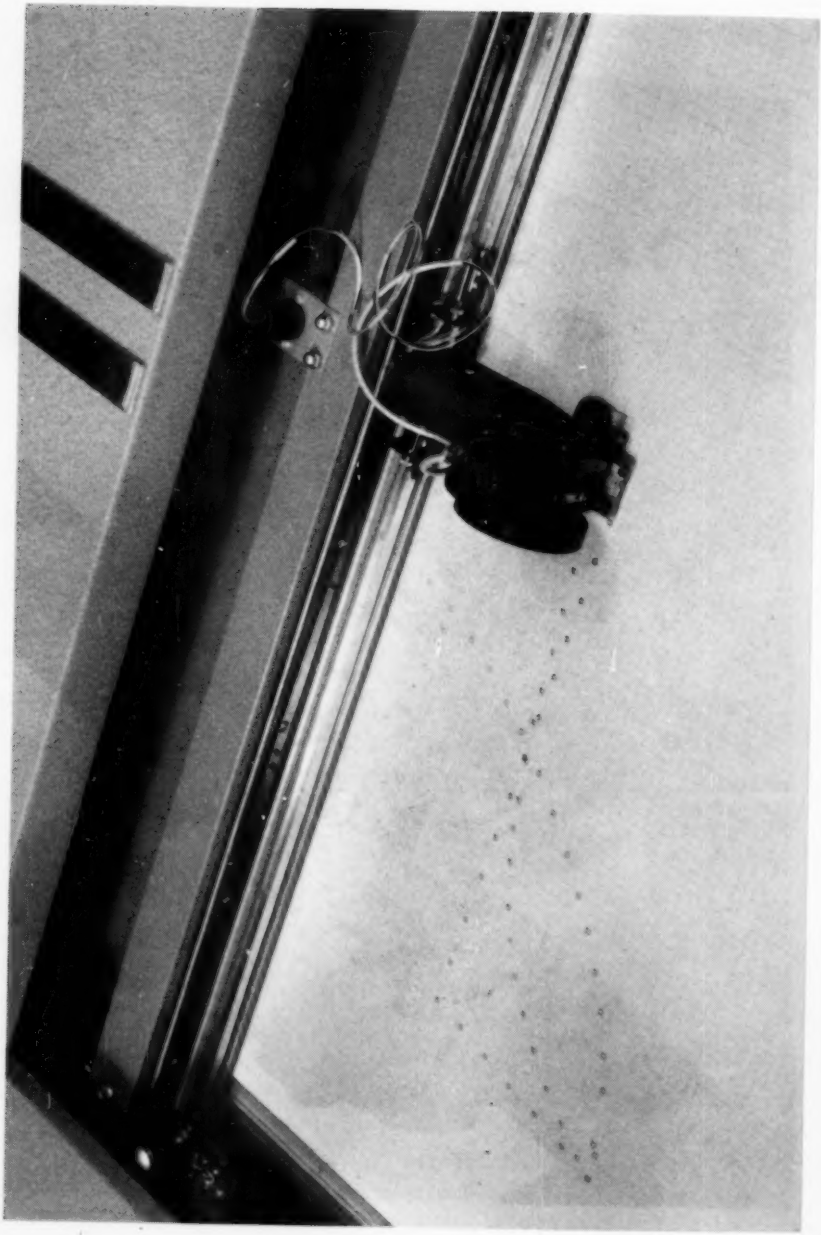
The speed of the Dataplotter, working from punched tape or punched cards, is about 50 points per minute depending somewhat on the

distance between points. Thus a deck of 2500 cards would take an hour to point plot on the Dataplotter. For line plotting, this speed is about 25 line segments per minute. A chart of some variable versus time in line plotting containing 120 points takes about 4½-5 minutes to draw. The magnetic tape unit, by taking advantage of faster reading time and other time saving techniques, plots discrete points at 6-8 points per second and about 10 line segments per second within the same accuracy as mentioned above. The magnetic tape plotter provides a free run mode which allows 75 point per second plot at .1" accuracy. This is a requirement frequently requested by users of magnetic tape output computers.

The Dataplotter has an accuracy of four digits. This means that it divides, electronically, the plotting surface into 10,000 counts ("0" - 9999). Thus, if all the data is positive and the origin is located at the left side of the board, the board would see the extreme right of the paper as 9999. Data of any value between these two positions would be plotted in a position on the board relative to these two extremes. To facilitate flexibility of charting, provision has been made to move the origin position off the board. This facilitates the expansion of the scale of the plot. By means of this off-board feature, we may remove the origin any amount up to 9,000 of the 10,000 counts from the plotting surface. At maximum, the plotter would be dividing the board into 1000 counts full scale, thus effecting an expansion of the scale 10 times. These adjustments allow charts or graphs of almost any scale in almost any position on the plotting surface to be plotted. The size of the form on which it is made may be any size up to 30" x 30" on our standard unit.

I trust that this short description has given you some idea of how the Dataplotter operates. The Dataplotter and plotters in general, are quite new. The use of plotting equipment, in conjunction with computers, can be a powerful technique. Much still remains to be done in the field of educating computer users. It is hoped that you gentlemen will consider this equipment for your own particular uses.





THE CHALLENGE OF RELIABILITY

by

W. A. MacCrehan
Manager of Quality
Aviation Products
Bendix Radio Division
Bendix Aviation Corporation

This growth in stature became a recognized fact only when top management recognized a return on quality investment - - returns aided in part by a specific demand on the part of military contracts to include the requirement that the contractor demonstrate a competency to control the manufacture of the product. This control included many areas broader than sampling plans, control charts or process capability studies. It included such things as procedures requiring detailed analysis of repetitive defects and predictability of an end equipment failure.

The philosophy of Quality Control, as it broadened, became more heavily directed toward prevention of the defect rather than the post mortem of corrective action once the defect had occurred - - yet the daily task of line production quality problems continued to require ingenuity for solution, and such techniques as diagnosing with diagrams, lot plot, span plan, to name a few, provided progress as Quality Control became solidly entrenched as a management tool.

Along with growth of operational Quality Control, came a broadening of the base on which statistical tools could provide an assist, such as inventory control through sampling, production linearity through operations research, sales forecasting - - - how many such programs were coordinated with the Quality Control Dept. could be measured in part on how professionally active was the Quality Control Dept. In a number of smaller companies, Quality Control and Industrial Engineering teamed up for numerous worthwhile studies leading to assembly cost reductions, scrap analysis, and defect per solder joint studies. The eagerness for Quality Control to prove that statistical methods for analysis was a useful tool proved to be

of tremendous value to the military in estimating the failure rate of carburetors, starters, generators for logistic purposes in stocking spare parts.

And now, we are here in the Missile Age - - - with this complexity has grown a responsibility far beyond the humble beginning of either inspecting or testing a product - - a responsibility involving process control and product performance predictability.

In its simplest form, Quality of the product is that measure of the product as received by the customer that can be inspected, tested, and judged so as to be acceptable or rejectable to a specification. The reliability of the product is that performance characteristic while in service as measured by mean time failure, removal per 1000 flying hours or some similar index of daily performance.

We, then, must establish a means of attaining a "know how" to achieve this additional measurable characteristic - - recognizing that responsibilities vary from company to company. The following list attempts to segregate the operations of traditional inspection and test, quality control, and the term "reliability assurance", which embrace the operational phases of a Total Quality Control approach.

INSPECTION & TEST DEPT.	QUALITY CONTROL	RELIABILITY ASSURANCE
Receiving Inspection	Prototype	Design Review
Line Inspection	Eng. Inspection	Component Evaluation
Line Test	Vendor Rating	Eng. Inspection
Final Inspection	Receiving Insp.	Vendor Rating
Final Test	Line Inspection	Receiving Inspection
Customer Complaint	Line Test	Value Analysis
	Defect Analysis	Line Inspection

QUALITY CONTROL	RELIABILITY ASSURANCE
Final Inspection	Line Test
Final Test	Quality Audit
Quality Audit	Test Equipment Analysis
Customer Complaint	Defect Analysis
	Final Inspection
	Final Test
	Quality Audit
	Customer Complaint
	Life Test & Type Test
	Product Predictability

It can be seen that if the professional Quality Control Dept. is to meet the challenge of reliability, it must recognize new areas of responsibility.

1. Design Review - This is close to the responsibility of the pure designer, but Quality Control files are packed solid with traditional testing data on component performance. Quality Control, to be an active partner with Engineering on design review, must know "mean time failure to impressed loads", derating effect on components, and above all, environmental characteristics of heat, cold, humidity, etc.

2. Engineering Prototype Inspection

This area couples the past experience of similar models with prediction of performance from new ideas - - an ideal area for the Quality Control Engineer to draw on his background of prevent failure both by structural considerations as well as prevention of defect failure where workmanship is a contributing

hazard.

3. Type Test, Life Test, Environmental Lab.

Much is known about room temperature operation; little is known of operating conditions where changes in the environment may contribute to product failure. The Quality Control Engineer is ideally adapted to designing experimental testing to randomize variables, to determining significant changes of confidence levels in testing results, to following and establishing controls for manufacturing processes that may be contributing to failure.

4. "Cradle to Grave" Product History

More and more significant is the requirement to know products by batch, lot, run or some similar identity. This, in turn, requires by date coding, serial number, color band or similar identity, an identification of the key components making up the assembly. This historical pedigree information allows prevention, prediction, or confinement where known failure cause is prevalent. The Quality Engineer has "cut his eye teeth" on data feedback - - of late, punch carding this information lets the Quality Control Engineer cross analyze by automatic sorting, tabulating and printing machines.

Many other areas could be listed where the Quality Control Engineer has "ready made training" for the reliability era, but what is going to tip the scales to make the Quality Control Engineer a vital factor in meeting the higher requirements of a dramatic new age is an awareness that professional Quality Control is broad - - - that a professional Quality Control Dept. is flexible, and above all, that a professional Quality Control Dept. is constantly progressive in

stepping up to meet the challenges posed by products specifically requiring conformance to a reliability clause contract.

We have known for some time how to make things good --- we fall short on being specific as to "how good", and until we can specify through "mean time failure", or some other such index, a performance characteristic, we have no quantitative expression of "how good". We are rapidly arriving at a point in professional Quality Control to where we can take a foremost position in organizing, planning and managing the function to quantitatively evaluate "how good", and reaching this plateau, a broad based, well organized, professional Quality Control Dept. can readily meet the challenge of reliability.